INTERNATIONAL TRADE
Theory and Evidence

James R. Markusen
University of Colorado, Boulder

James R. Melvin
University of Waterloo

William H. Kaempfer
University of Colorado, Boulder

Keith E. Maskus
University of Colorado, Boulder

McGraw-Hill, Inc.
New York St. Louis San Francisco Auckland Bogotá Caracas
Lisbon London Madrid Mexico City Milan Montreal
New Delhi San Juan Singapore Sydney Tokyo Toronto

CHAPTER 12

INCREASING RETURNS TO SCALE

12.1 INTRODUCTION

It has long been recognized that economies of scale provide an opportunity for trade and gains from trade. Yet very little theoretical research has been done on scale economies until recently. This is in spite of the fact that empirical research in countries as large as Canada and the Western European nations has emphasized the benefits when trade allows domestic firms to rationalize production. High protective tariffs in some of these countries have encouraged the manufacturing industries to produce a wide range of goods in small production runs rather than concentrating on larger and more efficient production of a narrower range of commodities. The more recent view is that even countries as large as the United States have not exhausted scale economies in the domestic market in some industries. Commercial aircraft and computers are two examples. United States consumers benefit from lower prices and higher quality in these industries because companies like Boeing and Intel can spread very high fixed costs over foreign as well as domestic sales.

The gains from trade due to scale economies can be understood fairly intuitively. Suppose, for example, that the cities and towns across the United States could not trade with one another. What do you suppose automobiles would cost if each town had to produce its own? The disadvantages of small-scale production would likely make automobiles prohibitively expensive in all but the largest cities. Trade allows production to be concentrated in a few large factories that ship to all parts of the nation. In fact, even a country

as large as Canada is too small to fully reap the scale economies inherent in automobile production. Canada entered into the Auto Pact of 1965 with the United States in order to achieve further scale economies. As a result of the increased trade permitted by the pact, plants in both countries (but especially Canada) produced fewer varieties of cars and increased the length of production runs, thereby lowering costs.

Scale economies are generally difficult to incorporate into general equilibrium models. They are usually inconsistent with perfect competition in production, for reasons discussed in the next section. Therefore, some of the techniques used in earlier chapters to analyze trade are no longer useful; a more complex analysis involving imperfect competition must be used.

12.2 EXTERNAL ECONOMIES

One exception to this problem of analyzing scale economies in general equilibrium occurs when they are external to individual firms, occuring instead at the industry level or at the level of industry groups (e.g., the manufacturing sector). Individual firms remain "small" so that the tools of competitive general equilibrium theory can still be used. It should be emphasized that this is not just a technical convenience. In the very first formal analysis of international trade, Adam Smith emphasized that increasing the size of the market permits a greater degree of specialization and therefore a higher level of productivity. It was only some time later that Ricardo introduced the notion of comparative advantage that became the cornerstone of international trade theory. Adam Smith's view was largely forgotten until recently.

Take agriculture, for example. Economies of scale for individual farms run out at a very small level of production, relative to the size of the market. Yet, as the size of the agricultural sector has grown, it has become profitable to produce specialized machinery and fertilizer, build railroads and handling facilities, conduct research into better seed varieties, and so forth. These developments have led to increasing production at decreasing cost while the industry as a whole captures scale economies, even though each individual producer may have constant returns to scale and may face fixed prices.

Let us return to the no-trade model and assume that all conditions of Chapter 6 hold, except that there are increasing returns to scale in one or both sectors. Suppose we have two absolutely identical economies whose production frontiers are both given by TT' in Fig. 12.1. The production frontier is bowed in (the production set is non-convex), under the assumption (as discussed in Chapter 2) that scale economies in X and/or Y outweigh any factor-intensity effects. Assume that the autarky consumption and production point for both countries is given by point A in Fig. 12.1.[1] It is interesting and important to note that in this case, no pattern of comparative advantage exists (autarky prices are equal), and yet there are potential gains to be had from specialization and trade. Suppose that Country H specializes

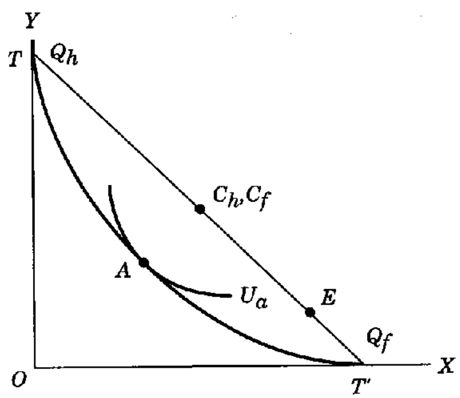


FIGURE 12.1
External economies.

in Y and Country F specializes in X, so that they produce at Q_h and Q_f, respectively, in Fig. 12.1. Each country can now trade half of its output for half of the other country's output, so that each country will consume at $C_h = C_f$. The utility level in each country will increase. Both countries will thus be better off, even though they were absolutely identical to start with. It is in this sense that scale economies offer gains from specialization above and beyond those obtained from any pattern of comparative advantage.

Although the consumption point $C_f = C_h$ in Fig. 12.1 could be an equilibrium, this is unlikely. An indifference curve for each country could be tangent to the cord TT' exactly at the midpoint, but only by chance.

Suppose that when the countries in Fig. 12.1 specialize, each wishes to consume at point E. This cannot be a trading equilibrium because the desired exports and imports of Country H exceed the desired exports and imports of Country F. In order to balance trade, the price of X will have to rise and the price of Y will have to fall. An equilibrium will eventually be reached as in Fig. 12.2, where Country H consumes at C_h and Country F consumes at C_f. The gains from trade are shared unequally even though the countries are identical. In fact, it is just possible that in equilibrium the price of Y is so low that the price line from Q_h passes below A, resulting in losses for Country H.

If Country H were a dictatorship, an all-knowing leader would naturally refuse to allow the country to specialize in Y in the case of Fig. 12.2. This could be accomplished by restricting the sale of Y to keep its price higher, for example. But the problem is that in a highly decentralized economy with external economies or imperfect competition (discussed subsequently), there is no mechanism to ensure that the country will not inadvertently end up as bad as or worse off than Country H in Fig. 12.2. The equilibrium in Fig. 12.2 is stable. In Country H the relative price of X in terms of Y (the slope of the price line) is less than the relative cost of producing X in terms of

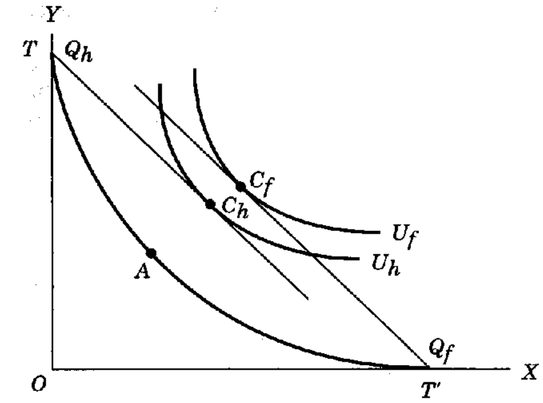


FIGURE 12.2
Uneven gains from trade.

Y (the slope of the production frontier). Thus H is in equilibrium specializing in Y. This demonstrates that with scale economies it is possible for the significant gains from trade to be very unevenly distributed and for one country to actually lose from trade.

This simple model of scale economies provides an interesting contrast to the Heckscher-Ohlin model with respect to the effects of trade on factor prices. Suppose the underlying production structure is Heckscher-Ohlin: there are two factors of production, and the factor intensities differ between industries. In the situation shown in Figs. 12.1 and 12.2, not only are factor prices not equalized by trade, but they are actually driven apart by trade. At the autarky equilibrium A in Fig. 12.1, the factor prices in the two identical countries must be equalized. With trade, one country specializes in Y and the other in X. Now because the two countries have the same technologies and factor endowments by the assumption that they are identical, each has the same total bundle of factors allocated to the good that it is producing. But if X is labor-intensive relative to Y, for example, the slope of an X isoquant through a given factor bundle must be steeper than the Y isoquant through that same factor bundle (e.g., refer back to Fig. 8.1). In each country, the factor price ratio w/r will be the slope of the isoquant of its good through the endowment point. Taken together, these arguments imply that if X is labor intensive, then the country specializing in X in Fig. 12.1 or Fig. 12.2 will have a higher w/r ratio than the country specializing in Y. The factor-price ratios, which are equal in the absence of trade, are driven apart by trade.

Two final points should be mentioned. First, note that this last example implies that each country in Figs. 12.1 and 12.2 has a relatively high price for the factor used intensively in its export industry. This has some very important implications for the effects of factor mobility, which we will discuss in a later chapter. Second, note that although *relative* factor prices

change with the introduction of trade, this does not imply that one factor is necessarily worse off with trade as it would in the case of the constant-returns Heckscher-Ohlin model. In the present case, the capture of scale economies in the good in which the country specializes means that the absolute productivities of both factors may rise at the same time that their relative productivities change. In the case in which both factors are better off (even if unequally so), there will be less opposition to trade liberalization.

12.3 INTERNAL ECONOMIES OF SCALE

We will now consider economies of scale that are internal to the individual firm. This applies to many manufacturing firms that obtain advantages from large-scale production. It is also often the case for some extractive industries such as mining and petroleum. Internal economies may also characterize some service industries such as banking, finance, and insurance.

The problem with internal economies of scale is that they are generally inconsistent with perfect competition and competitive equilibrium. The problem is examined in Fig. 12.3. Suppose a firm must incur some large fixed cost in plant or equipment in order to start production but can thereafter produce with constant returns to scale, or more precisely, at a constant marginal cost. Total cost will be given by

$$\mathrm{TC}_x = F + (\mathrm{MC}_x)X; \qquad \mathrm{AC}_x = F/X + \mathrm{MC}_x \tag{12.1}$$

where F is fixed cost, MC_x is marginal cost, and X is output. Average cost is equal to TC/X and is given by $\mathrm{AC}_x = F/X + \mathrm{MC}_x$. The situation is shown in Fig. 12.3 where MC_x is constant. AC_x falls steadily as the fixed cost is spread over a larger and larger output. AC_x approaches but never actually touches MC_x.

In this situation, could we have a competitive equilibrium in which price equals marginal cost? The answer is no, as we can show in a simple proof by contradiction. Suppose that the current market price is given by $p = p_c = \mathrm{MC}_x$ in Fig. 12.3, and that each firm perceives this price to be constant. At any output, each firm will lose money by virtue of the fact that $\mathrm{AC}_x > \mathrm{MC}_x = p_c$ or that the average cost exceeds price, no matter how much the firm produces. Therefore, price cannot equal marginal cost in equilibrium. If, for example, the firm produces at X_c in Fig. 12.3 where demand cuts the marginal cost curve, the firm's losses (negative profits) are given by the vertically hatched rectangle. This area is the per-unit loss of price minus the average cost $(p_c - \mathrm{AC}_c)$ times the number of units produced (X_c) or $(p_c - \mathrm{AC}_c)X_c$.

Suppose instead that the current price exceeds marginal cost and that each firm believes it can sell all it wants to at this price. In this case, each firm will attempt to produce an infinitely large output, because at some point AC will fall below price and then keep falling. Thus, no competitive equilibrium exists with internal increasing returns since either (1) the price is equal to (or below) marginal cost, in which case no firm will wish to

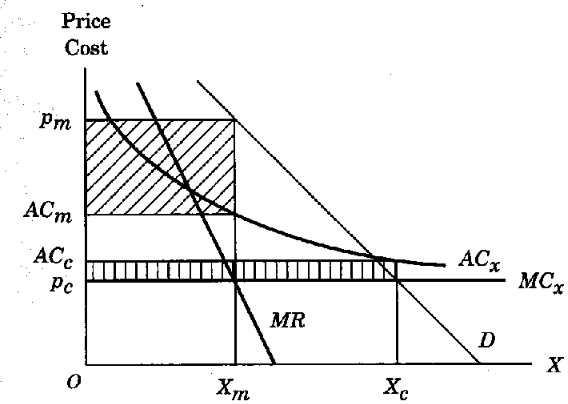


FIGURE 12.3
Partial equilibrium analysis of monopoly.

produce anything, or (2) the price is above marginal cost, in which case each firm will attempt to produce an infinite amount.

One feasible solution to this dilemma is for one large firm to monopolize the industry and to somehow prevent the entry of other firms (e.g., a second firm might calculate that if it enters, it will incur losses). We then have the standard monopoly outcome shown in Fig. 12.3 where MR is the marginal revenue curve drawn with respect to the demand curve D. Monopoly output and price are given by X_m and p_m, respectively, where MR = MC. Profits are given by the diagonally hatched rectangle, $(p_m - \mathrm{AC}_m)X_m$, or unit profit times output.

Alternatively, we could allow for free entry of firms until the demand curve facing each firm is driven down to tangency to the AC curve at one point. But in this and all possible equilibria we must be left with the situation that price exceeds marginal cost ($p >$ MC). This is the key result for our purposes: price must exceed MC or the firms will be losing money.

Figure 12.4 gives a general equilibrium representation of this fixed cost plus constant marginal-cost technology, which we will use for the remainder of the chapter. Suppose that X and Y are produced from a single factor, labor (L), which is in fixed supply ($\bar{L} = L_x + L_y$). Assume further that Y is produced with constant returns to scale by a competitive industry so that units can be chosen such that $Y = L_y$. If Y is chosen as *numeraire* ($p_y = 1$), the wage rate in terms of Y will also equal 1. p will denote the price of X in terms of Y, with the cost of producing X given by L_x.

It is assumed that the production of X requires an initial fixed cost, given by F, and then a constant marginal cost, MC_x, as in the preceding example. The total cost of labor required to produce X is then $L_x = F + (\mathrm{MC}_x)X$. The production frontier for this economy is shown in Fig. 12.4 as TFT'. $T = \bar{L}$ is the maximum output of Y when $X = 0$. To begin

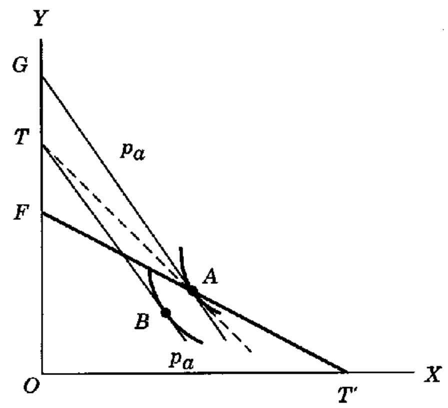


FIGURE 12.4
Monopoly equilibrium and the distribution of income.

producing X, the fixed cost TF must be invested before any output is realized. Thereafter, the constant marginal cost of producing X gives the linear segment FT', which has a slope equal to MC_x. This "kinked" production frontier is just a special case of the non-convex production set shown in Figs. 2.13 and 2.14 in Chapter 2. Note that although the production frontier in Fig. 12.4 is composed of linear segments, it has a property similar to those in Figs. 12.1 and 12.2 in that it lies everywhere below the diagonal connecting its endpoints T and T'.

The average cost of producing X is total cost (L_x) divided by output, or

$$AC_x = \frac{L_x}{X} = \frac{(\bar{L} - L_y)}{X} = \frac{(T - Y)}{X} \qquad (12.2)$$

Consider point A in Fig. 12.4. Equation (11.2) shows, for example, that the average cost of producing the amount of X at A is simply the slope of the line passing through T and A (the dashed line in Fig. 12.4). As we move along the linear segment FT', we see that the average cost of X is everywhere decreasing in the output of X, or alternatively, that production of X is characterized by increasing returns to scale. The equilibrium price ratio must cut the production frontier if positive X is produced at non-negative profits. This result is simply the general equilibrium representation of the result in Fig. 12.3 that price must exceed marginal cost. In Fig. 12.4, the slope of the price ratio must exceed the slope of the production frontier along FT', which is equal to MC_x, the marginal cost of production.

An autarky equilibrium with strictly positive profits for a monopoly producer of X is shown in Fig. 12.4, where the price ratio p_a through equilibrium point A is steeper than the average cost of X, given again by the slope of the dashed line TA. Point G in Fig. 12.4 gives the GNP of the country in terms of good Y. Workers are paid a wage of 1 in terms of Y (because the marginal product of labor in Y is 1), and so total wage income is given by $\bar{L} = \bar{Y}$ or OT in Fig. 12.4. GNP is thus composed of wage income in terms of $Y(OT)$ and profits in terms of $Y(TG)$. The budget line of wage earners is given by a line with slope p_a through $T = \bar{L}$ as shown in Fig. 12.4.

Point B on the wage earners' budget line represents the consumption budget of labor in Fig. 12.4. Therefore, the difference between the consumption bundles B and A is consumption out of profit income. Since wage earners' income is fixed at $T = \bar{L}$ in terms of Y, a decrease in p always increases wage earners' utility or real income (their budget line rotates around the fixed point T). Wage income equals GNP if profits are zero (e.g., if the price ratio is given by the dashed line TA in Fig. 12.4). Trade will change GNP through changes in workers' utilities and/or changes in monopoly profits. These two components of GNP may, of course, change in opposite directions.

12.4 SOURCES OF GAINS FROM TRADE WITH INCREASING RETURNS

Although there is always a *possibility* that a country could lose from trade when there are distortions, there are several conceptually separate sources of gains from trade in the presence of increasing returns and the associated imperfect competition. Which of these are captured depends on the given situation. For example, the pro-competitive effects of trade may lead to the exit of some firms and the consequent savings of fixed costs.

Pro-Competitive Gains

This was discussed in the previous chapter, so we will repeat the point only briefly here. Scale economies imply that the market can support only a limited number of firms, which will consequently be imperfectly competitive. Trade creates a larger market that can support a larger number of firms and the resulting greater level of competition. The pro-competitive effect of trade has been defined and measured in two ways in the economics literature. One way is to define it as a lowering of the markup on a firm's output. The other is to define it as the expansion of a firm's output (the product-expansion effect), with the resulting capture of price over marginal cost. There are technical inaccuracies in both definitions. Lowering the markup while holding the firm's output constant yields ambiguous welfare results. An expansion of output that captures the excess of price over marginal cost can occur in an external-economies model with no imperfect competition at all. These two effects are often inseparable. In the case of Cournot competition in Section 11.3 and Figure 11.4, the fall in the markup and the expansion of production occur together. *Both* are caused simultaneously by the perception of more elastic demand.

As in most of the literature, it is more convenient for our purposes to define the pro-competitive effect as identical to the product expansion effect of Chapter 11. When we do this, the pro-competitive effect can be decomposed into the sum of two separate components. Recall from the previous chapter that welfare is affected by a change in the output of X when X is priced above marginal cost by the amount

$$(p - MC_x)\Delta X \qquad (12.3)$$

An increase in the output of X is beneficial in that the economy captures the excess of price (the value of an additional unit in consumption) minus marginal cost (the value of the resources needed to produce an additional unit). Now write the total cost of producing X, TC_x, as average cost times output (since average cost is just total cost divided by output).

$$TC_x = X(AC_x) \qquad \Delta TC_x = AC_x \Delta X + X \Delta AC_x \tag{12.4}$$

Now divide the change in total cost in Eq. (12.4) by the change in output to get marginal cost.[2]

$$MC_x = \frac{\Delta TC_x}{\Delta X} = AC_x + X\left[\frac{\Delta AC_x}{\Delta X}\right] \tag{12.5}$$

Substitute the right-hand equation in Eq. (12.5) into Eq. (12.3) for MC_x.

$$(p - MC_x)\Delta X = (p - AC_x)\Delta X - X\left[\frac{\Delta AC_x}{\Delta X}\right]\Delta X \tag{12.6}$$

The pro-competitive (product-expansion) effect $(p - MC_x)\Delta X$ thus decomposes into two effects. The first term on the right-hand side of Eq. (12.6) can be called the *profit effect.* If price exceeds average cost (it need not always do so), then an increase in output generates a surplus of price over average cost on the additional output. This surplus is part of national income, although it may be very unevenly distributed. The second term can be called the *decreasing-average-cost* effect. With increasing returns, the change in average cost with respect to output is *negative*: $\Delta AC_x/\Delta X < 0$. Thus, when we include the minus sign in Eq. (12.6), we see that an increase in output improves welfare in that the average cost of producing the initial output X falls. When more output is produced, the initial output requires fewer resources.

The profit effect. We have just explained how a country can gain from trade when its firm expands output, capturing the excess of price over average cost on incremental output. Note that a gain here is not inconsistent with a fall in profit to zero as a consequence of trade, which is why the term *profit effect* is not a perfect one. Trade may lead the price of the product to fall so much that, in the new equilibrium, price is equal to average cost and profits are zero. But in the process of moving from autarky to free trade, the economy captures the excess of price over average cost (even if that difference is shrinking) on each additional unit of output produced. In a model with external economies of scale where price is equal to average cost for each firm, the profit effect is zero.

The decreasing-average-cost effect. The second effect in the decomposition of the pro-competitive (or product-expansion) effect derives from the fall in the average cost of producing the initial output when the firm expands its output. This is a savings in real resources, whereas the profit effect arises from the imperfect competition that generally accompanies scale economies. This effect can occur only with increasing-returns technologies. If we had

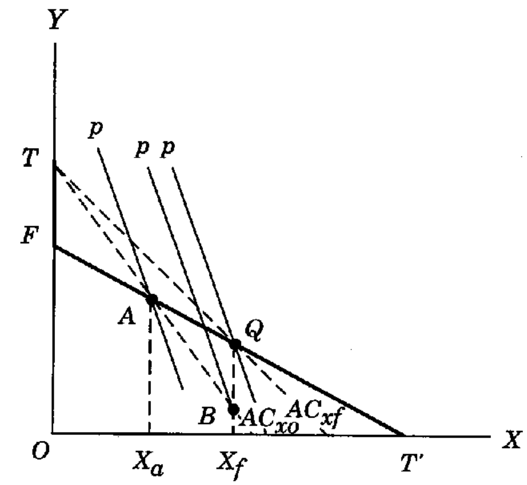


FIGURE 12.5
Decomposing the pro-competitive gain from trade.

imperfect competition with constant returns to scale, average and marginal cost would be the same, and we would have only the profit effect.

These two effects are separated in Fig. 12.5. Point A shows the initial equilibrium at price ratio p. Suppose that p stays the same with the introduction of trade (not very likely) but that the X producer expands output to point Q in Fig. 12.5. Now the country is clearly on a higher budget line. The movement from A to Q can be broken down into two components: A to B, and B to Q. First, consider the same expansion of X output from X_a to X_f, but holding average production costs constant at their initial level, AC_{x_0}. The latter is given by the slope of the dashed line through points T and A. Therefore, the movement from A to B is the profit effect as defined above. The movement from B to Q is the decreasing-average-cost effect. We can produce more Y at Q than at B because the average cost of production has fallen from AC_{x_0} to AC_{x_f}.

Figure 12.5 also shows two special cases we have mentioned. First, if the price ratio p equals the initial average cost AC_{x_0}, then there is no profit effect (the economy does not move to a higher budget line). The entire movement from A to Q would be the decreasing-average-cost effect. Second, if there were constant returns to scale (no fixed cost), then average costs would coincide with the slope of the production frontier and there would be no decreasing-average-cost effect. The entire effect of moving from A to Q would be due to the profit effect.

Figure 12.6 shows a possible outcome of trade in which two identical countries are combined, very much like the Cournot case of Fig. 11.4. The autarky equilibrium of both countries is at point A, with price ratio p_a. When trade is open, we assume that the two monopoly exporters behave according to the Cournot pricing rule discussed in Section 11.3 of the previous chapter.

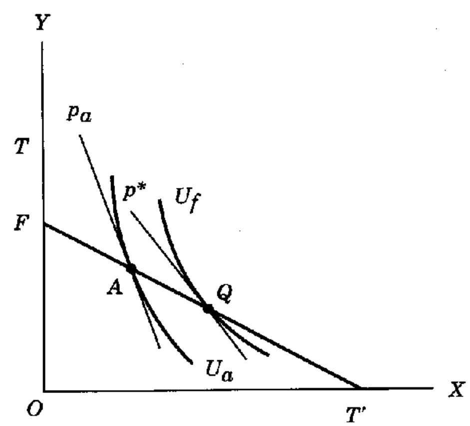


FIGURE 12.6
Cournot competition and pro-competitive gains from trade.

Each producer now views demand as more elastic and increases output. The final equilibrium could be at a point like Q at price ratio p^* in Fig. 12.6, with no net trade between the two identical countries but a gain from trade, nevertheless. Utility rises from U_a to U_f. This pro-competitive gain is due to both a profit effect and a decreasing-average-cost effect as defined above.

Firm Exit Effect

The gain from trade described in the preceding section resulted when the two existing firms increased their outputs in response to a perceived increase in the elasticity of demand. But what if the initial situation were characterized by the entry of firms up to the point where profits are zero? Then the increased competition induced by trade might lead to losses and to the exit of some firms.

Scale economies pose something of a dilemma with respect to the number of firms in an industry. On the one hand, it is desirable in terms of technical efficiency that there be a small number of firms. If the average-cost curve is everywhere downward-sloping (as in our current example), it is desirable to have the entire industry output produced by a single firm. But a small number of firms generally implies greater market power and correspondingly greater monopoly output restriction. This trade-off for a single economy gives rise to a third source of gains from trade. With trade we can increase the total number of firms in competition while reducing the number of firms in each individual country. For example, suppose two countries with 10 firms each enter into trade with each other and that the resulting competition drives out three firms in each country. The number of firms in each country has decreased, but the total number competing with one another is now 14, an increase over the 10 that were competing in each country in autarky.

More advanced theoretical treatments have shown that this is exactly the kind of outcome we will get if there is Cournot-Nash competition and free entry of firms in each country. Free entry drives profits down to zero in each country in autarky. But the opening of trade causes each firm to perceive demand as more elastic, as was discussed in the previous chapter. Generally, firms will increase output, and some firms will then exit as profits are initially negative. The trading equilibrium will have fewer firms in each country, with each firm producing a higher level of output at lower average cost. Empirical analyses for Canada have confirmed that trade liberalization has this rationalizing effect on manufacturing firms.

The situation is illustrated in Fig. 12.7, where A is the autarky equilibrium for each of two identical countries. Free entry has forced price down to average cost, and the vertical distance TF' is now interpreted as the combined fixed costs of the existing firms. Trade causes each Cournot-Nash firm to perceive its demand curve as more elastic and therefore to increase output. But this leads to negative profits and the exit of some firms. Equilibrium is restored at a lower price with fewer firms in each individual country but more firms in total. Exit of the redundant firms frees up the resources that had been devoted to fixed costs and shifts the production frontier of each country out to TFT''. No trade, or more correctly, no net trade, need occur, because both countries are identical and both attain consumption point C in Fig. 12.7.

The movement from A to C in Fig. 12.7 is a combination of a pro-competitive effect and an exit effect. We can decompose the movement from A to C by considering point B. At B, output per firm must be the same as at C, because the average cost of producing at C is the same as at B (recall from Eq. (12.1) that average cost falls continuously in firm output). Points A and B involve the same number of firms because the same resources are devoted to total fixed costs. Thus, the movement from A to B is a pure expansion in

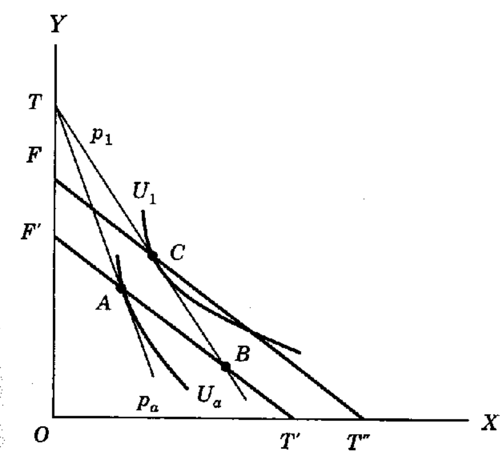


FIGURE 12.7
Exit of redundant firms.

firm outputs, holding the number of firms constant. A to B is therefore the pro-competitive gain from trade. The movement from B to C is a reduction in the number of firms, holding the outputs of surviving firms constant, and therefore shows the firm-exit effect.

Increased Product Diversity

A recent series of papers has emphasized increased product diversity as a form of gains from trade. This emphasis is usually referred to as the *monopolistic competition model* of trade because it draws on a model of the same name used in microeconomics and industrial organization theory. The idea is that an industry is "competitive" in that there are a large number of firms, but it is also "monopolistic" in that each firm is producing a somewhat unique product. The latter assumption implies that these firms do individually face downward-sloping demand curves.

The situation is shown in Fig. 12.8, where we now assume that both X and Y are produced with increasing returns to scale. Production functions for X and Y are identical, and the goods are symmetric but imperfect substitutes in consumption. This last assumption means that consumers are indifferent between one unit of X and one unit of Y, but that they would like to have some of both rather than more of just one. The consumer is, for example, indifferent between a stereo and a television, but would rather have one stereo and one television than either two stereos or two televisions. This model, which reflects consumer preference for diversity, is called the *love of variety* approach.

In autarky, each of the identical countries in Fig. 12.8 could attain consumption point A. But this is now not the best choice, even in autarky. Because of the large fixed costs, it is in a country's interest in autarky to produce only one good, specializing in either Y at point T or in X at point

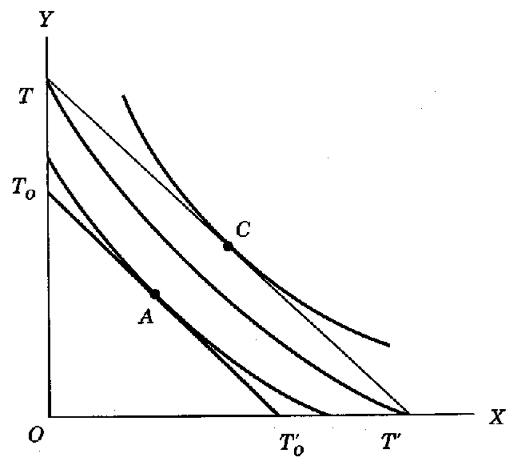


FIGURE 12.8
Increasing product diversity: "love of variety."

T' in Fig. 12.8. This puts the country on an indifference curve through T and T', which is higher than the indifference curve through A. The benefits of product diversity are outweighed by the high fixed costs of starting production of the second good. With trade, each country could specialize in one of the goods and trade half of its output for half of the output of the other country's good, thereby attaining consumption point C (again, this is unlikely to be the actual equilibrium). In this situation, there is no change in the average cost of a good, nor are there any procompetitive gains. Instead, where scale economies were limiting the number of goods consumed before trade, consumers "decided" to take the gains from trade in the form of having more products rather than having lower costs for their existing products.

Gains from trade through product differentiation can be looked at in a second way. For example, although consumers may buy only one automobile each, each would choose a different "ideal" automobile, depending on tastes and income level. This approach to product diversity is labeled the *ideal variety* approach. Because of scale economies, no country can afford to produce a unique automobile for each consumer. Germany produces Volkswagens and Mercedes-Benzes, and France produces Renaults and Peugeots, all of which have somewhat differing characteristics. Trade in automobiles occurs between France and Germany because some Germans prefer Renaults or Peugeots and some Frenchmen prefer Volkswagens or Mercedes-Benzes.

This situation is shown in Fig. 12.9. Suppose that automobiles have only two characteristics, size and fuel efficiency. There is a tradeoff between these two characteristics: buying a bigger car means sacrificing some fuel efficiency. Figure 12.9 shows three possible combinations of size and efficiency, denoted X, Y, and Z, corresponding to three different types of cars. Suppose that all three models could be produced at the same average cost for the same volume of production. But because of scale economies, the average cost rises steeply as sales fall. Assume finally that societies consist of only

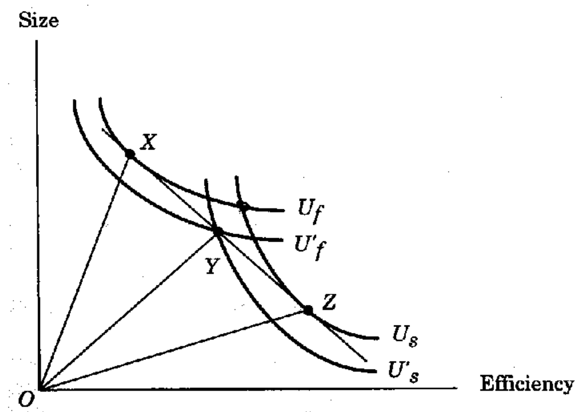


FIGURE 12.9
Increasing product diversity: "ideal variety."

two groups, students and faculty, and that the former have a relatively high preference for efficiency and the latter a relatively high preference for size. Their indifference curves are U_s and U_f in Fig. 12.9, respectively.

Even though the three models cost the same at the same level of production, scale economies force a choice. If the country produces both X and Y, giving both students and faculty their "ideal" cars, the volume of sales of each model will be much lower and the average cost much higher than if only the single compromise model Y were produced. Thus, an entirely possible outcome is that only model Y is produced. It sells for a modest cost, but faculty are cramped and students are poor from buying gas. In Fig. 12.9, these groups attain indifference curves U_f' and U_s', respectively. Now if we add a second, identical country through trade, one country could produce model X and the other model Z, each exporting half its output for half the foreign output of the other model. Each country would be producing the same number of cars as in autarky, and consequently, the cars would have the same average cost. Assuming X and Z sell for the same price that Y did in autarky, consumers pay the same for cars but attain the higher indifference curves U_f and U_s. As in the example of Fig. 12.8, consumers take the gains from trade in terms of increased product diversity rather than in the form of lower prices.

Specialized Plants and Inputs

Increases in market size due to trade may also permit firms to build plants specialized to fewer product lines and to create specialized inputs. Two of many possible examples are machinery that is highly specialized to certain products and consultants who specialize in certain business or engineering problems. Little theoretical work has been done on the former, although empirical analysis in Canada has emphasized plant specialization as an important source of gains from trade liberalization. The U.S.-Canada Auto Pact of 1965 was partly motivated by the fact that Canadian plants were producing too many models in small production runs.

Figure 12.8 can be used to depict the gains from specialized inputs with indifference curves now interpreted as isoquants for final output. Creation of specialized inputs requires an initial fixed cost of TT_0. In autarky, each country creates only one such input in the simple example of Fig. 12.8. An output level corresponding to the isoquant through T and T' is attained in each country from the amount of resources devoted to producing this good. When the countries are combined through trade, the creation of two specialized inputs is profitable because the output level at C can be attained if each country trades half of the output of its specialized input for half of the other country's product.

Before moving on, recall the point about factor returns discussed briefly at the end of Section 12.2. When there is more than one factor, scale economies, whether resulting in lower prices for existing goods or in more goods, make it more likely that all factors will gain from trade. While relative factor prices may change as in the Heckscher-Ohlin model, lower prices for existing goods relative to factor prices or the benefits of having a more diverse consumption bundle, may mean gains from all factors. In the case of increased product diversity, for example, the wage rate could fall in the production of one of the differentiated products, but this decrease in welfare could be outweighed by the utility gain that results from having more products. This effect generates even more political support for free trade.

12.5 AN EXTENSION OF THE COURNOT MODEL TO FREE ENTRY AND EXIT

Suppose we take the Cournot model of the previous chapter and add free entry and exit by firms until profits are zero. Now we will have two equilibrium conditions, $MR_x = MC_x$ as in Chapter 11, and the zero-profit condition $p_x = AC_x$. Suppose all firms are identical so that we can deal with a representative firm i. We can use our cost function from Eq. (12.1) to write the second condition as marginal cost plus F/X_i, where X_i is the output of the representative firm.

$$p_x(1 - s_i/e) = MC_x \qquad p_x = MC_x + F/X_i \tag{12.7}$$

where s_i is the market share of the representative firm. Divide the second equation by the first.

$$\frac{1}{1 - s_i/e} = 1 + \frac{F}{(MC_x)X_i} \tag{12.8}$$

Assume that fixed costs (F), marginal costs (MC_i), and the market elasticity of demand (e) are all constants. Now assume that we take two identical countries and combine them in trade. This is equivalent to doubling the size of a single country. Under the assumptions noted, only one result is possible: *s_i must fall and X_i must increase.* They cannot both stay constant; with the market growing, a constant output per firm must mean that the individual firm's market share is falling. s_i and X_i cannot increase or decrease together; that would mean that Eq. (12.8) is no longer satisfied. The equation can be satisfied if s_i increases and X_i falls, but this is impossible in the growing market; a firm's market share cannot grow in a growing market if its output is falling (this can be seen more formally by replacing s_i with X_i/X, where X is total sales).

We have now determined that the larger market induced by trade leads to an expansion in output per firm. This by itself is a sufficient condition for gains from trade. But we can carry the model a bit further. If all firms are identical as we have assumed, then the market share of a representative firm, s_i, is just $1/N$ where N is the number of firms. The fact that s_i falls when we combine two identical economies means that the total N must rise. But the number of firms cannot in the end be double (that is, each country cannot continue to support the same number of firms as in autarky) if we

place some simple restrictions on the form of demand functions.[3] Suppose that the number of firms did double (the number of firms in each country stayed constant). Because each firm increases its output due to an increased perceived marginal revenue (s_i falls), profits of the existing firms will fall below zero, leading some firms to exit. In this case, we get the result shown in Fig. 12.7. Trade leads to an exit of some firms in each country but leaves the total number of firms in the expanded economy larger than in either country in autarky. The market is more competitive, surviving firms have lower average costs and higher outputs per firm, and the two identical countries necessarily gain from trade.

12.6 AN ALGEBRAIC VERSION OF THE MONOPOLISTIC-COMPETITION MODEL*

In this section we will take a more formal look at increased product diversity in a monopolistic-competition model. The purpose of the section is to show more rigorously how the number of products is determined and to demonstrate that we can indeed construct a case in which all gains from trade are in the form of increased diversity. The model we will use corresponds very closely to the situation shown in Fig. 12.8. Suppose that the X sector is actually composed of many firms producing somewhat different products. Suppose that the utility functions of identical consumers are given by

$$U = \sum_{i=1}^{n} X_i^{\alpha}, \qquad 0 < \alpha < 1 \tag{12.9}$$

where the number of products produced, n, is *endogenously determined.* Each product that is produced yields diminishing marginal utility, and so the consumer would always rather have one each of two products, say X_1 and X_2, rather than two units of either. Labor is the only factor of production. We will assume free entry of firms such that profits are driven to zero. The economy's budget constraint is thus that labor income is spent on X.

$$\bar{L} = \sum_{i=1}^{n} p_{xi} X_i \tag{12.10}$$

More advanced treatments will show that the demand function for X_i is given by the following rather complicated formula.

$$X_i = \frac{\bar{L}}{p_{xi}^{\sigma}\left[\sum_j p_{xj}^{-\alpha\sigma}\right]} \qquad \text{where} \quad \sigma = \frac{1}{1-\alpha} \tag{12.11}$$

*The analysis in this section is considerably more difficult. You can skip it without losing continuity and refer to Fig. 12.8 for a grasp of the basic ideas behind the monopolistic-competition approach.

If there are many X producers, we can safely assume that each producer treats the term in square brackets as a constant. Students with calculus can find the elasticity of demand for X_i as p_{xi}/X_i times the derivative of Eq. (12.11) with respect to p_{xi}. The elasticity of demand is given by

$$e_i = -\frac{dX_i/X_i}{dp_{xi}/p_{xi}} = \sigma = \frac{1}{1-\alpha} \tag{12.12}$$

The X producers have the same technology that we described in Eq. (12.1). Let labor be numeraire, so that $w = 1$. p_{xi} is the price of X_i in terms of labor. Producer maximization then requires that the firm set marginal revenue equal to marginal cost, MC_x.

$$p_{xi}\left[1 - \frac{1}{e_i}\right] = p_{xi}[1-(1-\alpha)] = p_{xi}\,\alpha = \mathrm{MC}_x \tag{12.13}$$

Freedom of entry and exit requires that equilibrium is characterized by price equals average cost.

$$p_{xi} = \mathrm{MC}_x + F/X_i \tag{12.14}$$

If we combine Eqs. (12.13) and (12.14), we can solve out for the equilibrium level of X_i, which is the same for all X_i that are produced (i.e., all X_i that are produced are produced in the same amount).

$$X = \frac{\alpha F}{\mathrm{MC}_x(1-\alpha)} \tag{12.15}$$

Goods are produced in a fixed amount independent of other factors such as the size of the economy, measured by the labor endowment. So suppose that we double the labor endowment, which we can again think of as combining two identical economies. The labor endowment must equal the total labor used in the n goods that are produced.

$$\bar{L} = n(MC_x X_i + F) = n\left[\frac{\alpha F}{1-\alpha} + F\right] = n\left[\frac{F}{1-\alpha}\right] \tag{12.16}$$

With goods produced in fixed amounts, the economy expands entirely by expanding the number of goods (an increase in $\bar{L}$ increases n in the same proportion). This is the situation shown in Fig. 12.8. Doubling the size of the economy through trade allows each consumer to purchase a more diverse consumption bundle, thereby capturing the welfare benefit from increased diversity shown in Fig. 12.8. Trade does not increase the scale of production of any good, yet it is scale economies that are responsible for the gain from trade. Scale economies are what limit the number of goods produced (degree of product diversity) initially.

Since any good continues to be produced in the same amount, as shown in Eq. (12.15), we can denote the production of a representative good simply as X. The utility function in Eq. (12.9) can then be written as

$$U = nX^{\alpha} \tag{12.17}$$

If we combine two identical economies, and if the output of any particular good remains fixed, consumers in each country will have twice as many

goods from which to choose and will consume half as much of each good as they did in autarky (the fixed amount of a good must now be shared among consumers in both countries). Treating n and X as the autarky values, denote free trade utility by U^f and autarky utility by U^a.

$$U^f = (2n)(X/2)^a = 2^{1-a}nX^a > nX^a = U^a \qquad \text{given that} \quad 2^{1-a} > 1 \tag{12.18}$$

Consumers gain through increased diversity. As suggested earlier, it is better to have one stereo and one television than to have two of either.

12.7 CONCLUDING REMARKS

Our analysis of scale economies has produced a number of results that differ in very basic ways from those of the more traditional Ricardian and Heckscher-Ohlin models. First, there is the question of the direction of trade. In the traditional models, the direction of trade is related to underlying characteristics of the economy. In the Heckscher-Ohlin model, for example, we saw that a country exported the goods that intensively used the country's relatively abundant factors. With scale economies, however, there may be some inherent arbitrariness in the pattern of specialization. This was illustrated in Fig. 12.1 and in subsequent diagrams, where we noted that with scale economies, trade can arise between two identical economies, or rather, between two economies for which there exists no pattern of comparative advantage. Of course, in reality the gains from scale economies occur in *addition* to gains due to comparative advantage. The methodology of the no-trade model is not to present *alternative* theories but rather to focus on individual determinants of trade and gains from trade one at a time. The theories are in fact *complementary* in that generally all of them have simultaneous and important roles in actual trade.

Countries typically specialize when there are increasing returns, but the welfare effects of trade may depend on which country specializes in which good. One implication of this result is that the observed pattern of trade in goods produced with returns to scale may often be determined by historical factors, such as which country entered an industry first. The first entrant gains an advantage (sometimes called *first mover advantage*) by achieving low costs and perhaps technical expertise early; this may discourage foreign entrants who might have to enter at an initial high cost.

Second, returns to scale have implications for the gains from trade, implications that differ in a number of ways from the more traditional results. On the one hand, scale economies offer more possibilities for gains, whereas on the other hand, there are possible complicating factors. As we showed earlier, scale economies may mean that there will be gains from trade even for two identical economies. Alternatively, gains from trade exist independent of any pattern of comparative advantage. The lower costs and increased product diversity achieved through the capture of scale economies are thus an additional source of gains that complement the gains achieved from other bases for trade, such as differences in factor endowments.

Strong conclusions about gains from trade are, however, impossible to draw because scale economies are generally associated with distortions, for reasons noted earlier. The result is that a country's prices may not accurately reflect the underlying costs and the pattern of comparative advantage. When this happens, a country can end up specializing in the wrong goods to the point of experiencing welfare losses from trade.

Many economists are finding empirical evidence that scale economies are a very important determinant of trade and the gains from trade for countries the size of Canada and many of the Western European nations. Empirical estimates of the gains from trade liberalization in both of these regions suggest that an important source of gains has been a rationalization of the manufacturing and certain service industries. Some would assert that these rationalization effects constitute a major source of gains from the formation of the EEC, outweighing factors having to do with comparative advantage. There is also an increasing acceptance of the view that issues involving scale economies and imperfect competition are important for a country as large as the United States. Finally, with the increased mobility of factors of production (especially capital) and technology across international borders, there is some reason to suggest that the Heckscher-Ohlin and Ricardian bases for trade may be relatively less important in explaining trade among the industrialized countries than they were even a few decades ago. The very large volume of trade among the very similar economies of North America, the EEC, and Japan seems to constitute prima facie evidence that scale economies, imperfect competition, and product differentiation are important determinants of trade.

The principal points of the chapter are summarized as follows.

1. With increasing-returns-to-scale technologies, trade and gains from trade can arise even between two identical economies. We could refer to this as *non-comparative-advantage trade* in contrast to trade generated by underlying differences among countries. We also showed that the gains from trade are not necessarily distributed evenly, even between identical countries.

2. There are several sources of gains from trade in the presence of scale economies, although they are often hard to disentangle in practice. Expansion of an increasing returns sector through trade generates a benefit equal to the excess of price over marginal cost on incremental output. This pro-competitive gain can be broken down into the excess of price over average cost (the profit effect) and the fall in the average cost of producing existing output (the decreasing-average-cost effect). Gains may also be captured as a result of the exit of some firms, which frees up the resources that were used in fixed costs for other uses.

3. Gains from trade may be captured from increased product or input diversity instead of from lower average costs for a fixed range of products. In other words, gains may be captured by having either the same range of products at lower costs or a larger range of products at the same costs.

Of course, these two extremes will generally be mixed in practice. The final sections of the chapter modeled the two extreme cases. Section 12.5 gave an extension of the Cournot model of the previous chapter, showing that combining two identical economies leads surviving firms to increase production (resulting in a decrease in the average production cost) of a fixed range of products. Section 12.6 used a version of the monopolistic-competition model, in which any good that is produced is produced in a fixed amount, to show that trade between two identical countries leads to gains from increased product diversity.

4. The chapter focused largely on a one-factor model in order to concentrate on the essence of scale economies. However, two points that are important in a multi-factor context were briefly noted. First, trade based on scale economies between similar economies may drive factor prices farther apart in the two countries. We will explore this further in Chapter 21. Second, scale economies make it more likely that all factors will gain from trade, contrary to the Heckscher-Ohlin model. While trade may change *relative* factor prices, the increased factor productivities due to the capture of scale economies may mean that the *absolute* or real returns to all factors rise. In the case of increased product diversity, a factor's return may fall in terms of a representative good, but this negative effect may be outweighed by the availability of a more diverse consumption bundle.

PROBLEMS

1. Show that if both countries want to consume at E in Fig. 12.1, there will be excess demand for X and excess supply of Y.
2. Show by redrawing Fig. 12.2 that the relative price of Y could drop so low that Country H would lose from trade.
3. Show that if a firm's price is less than its average cost, its profits must be negative.
4. What is the largest output that the firm could produce in Fig. 12.3 and still break even?
5. Using Fig. 12.7, show that point C could not be attained by the two countries if some firms did not exit from the industry.
6. Redraw Fig. 12.8 with different indifference curves such that autarky equilibrium at A is preferred to an equilibrium at either T or T'.
7. In the discussion of product diversity and monopolistic competition, we consider two different situations: (a) consumers have identical tastes but prefer diversity (*love of variety*, Fig. 11.8) and (b) consumers have heterogeneous tastes (*ideal variety*, Fig. 11.9). Determine which approach is likely to be more useful in each of the following situations: (1) choosing groceries in a supermarket, (2) buying a house, (3) shopping for clothing, (4) choosing a new stereo.

NOTES

1. Equilibrium is generally not given by a tangency between an indifference curve and the production frontier. The slope of the production frontier incorporates the externality effects that one producer has on another, while private firms do not take these effects into consideration in making their input and output decisions. If, for example, there are external economies in the X sector, then the "social" marginal cost of producing an additional unit of X is less than the private marginal cost, which is equal to the price of X, p_x. In equilibrium, we would have $p = p_x/p_y >$ MRT. The price line would cut the production frontier as in Fig. 11.1 of the previous chapter. Figure 12.1 could be generated by an economy in which the external economies are equally strong in both sectors (not very likely, but it helps make the simple point we wish to show).
2. Students with calculus will recognize the second equation in Eq. (12.4) as the rule for differentiation of a product. Strictly speaking, it is valid only when Δ approaches zero.
3. A sufficient condition is that demand is not "too convex." A linear demand curve is sufficient for the result that some firms in each country must exit. The difficulty is that, if demand is very convex, the price fall caused by the pro-competitive effect causes consumers to spend much more on X in total. Holding the initial number of firms at the combined autarky total, if this effect is sufficiently strong, the markup revenue received by a firm may increase even though that firm is cutting its markup *rate* in half. Basically, if combining the two economies does not at least double revenues for each firm (causing markup revenues to fall), then some firms will exit the market.

REFERENCES

Ethier, W. (1979). "Internationally Decreasing Costs and World Trade." *Journal of International Economics* 9: 1–24.
Ethier, W. (1982). "National and International Returns to Scale in the Modern Theory of International Trade." *American Economic Review* 72: 389–406.
Grossman, G. (1992). *Imperfect Competition and International Trade.* Cambridge: MIT Press.
Helpman, E. (1981). "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition. A Chamberlinian, Heckscher-Ohlin Approach." *Journal of International Economics* II: 304–340.
Helpman, E., and Krugman, P. (1985). *Market Structure and Foreign Trade.* Cambridge: MIT Press.
Horstmann, I., and Markusen, J. R. (1986). "Up the Average Cost Curve: Inefficient Entry and the New Protectionism." *Journal of International Economics* 20: 225–248.
Harris, R. G., and Cox, D. (1984a). *Trade, Industrial Policy, and Canadian Manufacturing.* Toronto: University of Toronto Press.
Harris, R. G. (1984b). "Applied General Equilibrium Analysis of Small Open Economies with Scale Economies and Imperfect Competition." *American Economic Review* 74: 1016–1032.
Kemp, M. C. (1969). *The Pure Theory of International Trade and Investment.* Englewood Cliffs, N. J.: Prentice-Hall.
Krugman, P. (1979). "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9: 469–479.
———. (1981). "Intraindustry Specialization and the Gains from Trade." *Journal of Political Economy* 89: 469–479.
Lancaster, K. (1980). "Intra-Industry Trade Under Perfect Monopolistic Competition." *Journal of International Economics* 10: 151–175.
Markusen, J. R. (1981). "Trade and the Gains from Trade with Imperfect Competition." *Journal of International Economics* 11: 531–551.
———. (1989). "Trade in Producer Services and in Other Specialized Intermediate Inputs." *American Economic Review* 79: 85–95.
Melvin, J. R. (1969). "Increasing Returns to Scale as a Determinant of Trade." *Canadian Journal of Economics* 2: 389–402.
Wonnacott, R. J., and Wonnacott, P. (1967). *Free Trade Between the United States and Canada.* Cambridge, Mass.: Harvard University Press.

CHAPTER 13

TASTES, PER CAPITA INCOME, AND TECHNOLOGICAL CHANGE AS DETERMINANTS OF TRADE

13.1 INTRODUCTION

Our models to this point have focused on the production side of the general equilibrium structure of economies to describe the causes of trade and the sources of gains from trade. This focus is representative of trade theory in general, which has devoted the overwhelming share of its attention to production, neglecting consumption almost entirely. Many models assume that consumers have identical and homogeneous utility functions, regardless of where they are located. This means that if commodity prices were equalized by trade, consumers everywhere would demand goods in the same proportions. All trade would then be due to various differences in production among countries. As students have realized by now, this emphasis generates a markedly diverse and sophisticated set of predictions about the supply-side determinants of comparative advantage, such as differences in factor endowments, policy-based market distortions, and imperfect competition associated with increasing returns to scale.

There has been much less theoretical attention paid to demand-based determinants of trade. Neither has there been much empirical examination of the apparent presumption by trade economists that production differences are more important than consumption differences for explaining trade patterns.[1] This phenomenon is curious in that the assumption of identical and homogeneous preferences across countries is easily shown to be questionable. For example, in 1991, 59 percent of household consumption in Bangladesh was estimated to have been devoted to food. The corresponding shares of food in household budgets in other countries were: Indonesia 48 percent, Argentina 35 percent, Greece 30 percent, Japan 17 percent, and United States 10 percent.[2] In contrast, budget shares allocated to medical care were: Bangladesh 2 percent, Indonesia 2 percent, Argentina 4 percent, Greece 6 percent, Japan 10 percent, and United States 14 percent. In part, these differences reflect variations in relative prices due to trade barriers and government policies. However, they clearly indicate that preferences are not homogeneous, in that the relative share spent on food declines and that of medical care rises as the level of economic development rises. They might suggest as well that tastes are also not identical, though this cannot be conclusively inferred from share data alone. However, an obvious national difference that would affect preferences is that the industrialized countries tend to have larger percentages of their populations at advanced ages, increasing the demand for health care. Accordingly, one purpose of this chapter is to bring together some strands of literature in order to examine the possible ways in which international differences in consumption patterns can influence trade flows.

A second shortcoming of our production-based models to this point is that they are entirely static in nature. They investigate the effects of differences in important supply characteristics on trade flows. However, most people will recognize that international trade bears important relationships to dynamic changes in the world economy, such as shifts in technology and new product innovation. These models are also not as fully developed as the standard trade theories, but it is important to bring them into our discussion of trade determinants. Thus, the second purpose of this chapter is to analyze some relatively simple aspects of the influence of technological change on international trade.[3]

We should place these newer theories in some historical context. Interest in developing trade models that could serve as alternatives to the Ricardian and Heckscher-Ohlin concepts emerged largely in the 1960s and 1970s after several empirical studies raised questions about the practical validity of these traditional models. We review these studies in the next chapter. However, most models based on demand variations, scale economies, and technological differences are relatively recent and are still under active theoretical development. This helps explain why there are many disparate views of what determines trade in these terms and why the models vary in complexity and rigor. However, these influences on trade are clearly important and deserve separate treatment here.

13.2 INTERNATIONAL CONSUMPTION BEHAVIOR AND TRADE

In this section we analyze the implications of allowing consumption patterns to vary between countries in two ways. First, we take national utility functions to be homogeneous but not identical. Second, we consider the case of identical but nonhomogeneous preferences.

Different Tastes

We begin with the most simple specification of international differences in preferences. Following our previous methodology, we assume that the only difference that exists between two countries is in demand conditions, for only in this way can we be sure that the results derived depend entirely on differences in demand. In particular, we assume that endowments in the two countries are identical, that production functions are identical between countries, and that production takes place with constant returns and perfect competition. These assumptions imply that the production possibility curves for the two countries will be identical. Thus, in Fig. 13.1, TT' represents the production possibility curve for both Country H and Country F. In this section we also assume that there are no distortions in either economy and that the utility functions of the two countries, while differing, are both homogeneous. The case of nonhomogeneous demand is considered in Section 13.3. The autarky equilibrium is the point where the highest indifference curve for each country is tangent to the production possibility curve. If we assume that tastes in Country H are biased toward commodity Y relative to tastes in Country F, then U_h and U_f could be representative indifference curves for the utility functions of H and F, respectively.

The autarky positions for countries H and F are A_h and A_f, respectively, and the autarky price lines are p_h and p_f in Fig. 13.1. Therefore, in autarky,

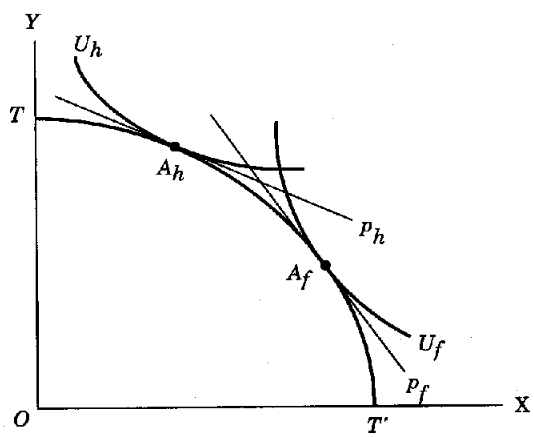


FIGURE 13.1
International differences in tastes.

commodity Y is relatively expensive in Country H, whereas commodity X is relatively expensive in Country F. The reason for this outcome is simply that have a stronger preference for good Y, driving up its price in comparison with X, with the opposite situation prevailing in F. When trade is permitted, the residents of Country H, observing that Y is relatively less expensive in the foreign country, will shift their purchases from the home country to Country F. Similarly, residents of the foreign country, observing that commodity X can be purchased more cheaply in Country H, will shift some of their purchases to that country. The results of these shifts in demand will be that the production point for the home country will move down the production possibility curve, whereas the production point for the foreign country will move up toward the Y axis. These adjustments will continue until there is no longer any incentive for residents of one country to increase their purchases in the other, or in other words, until commodity trade has succeeded in equalizing commodity prices. Such a situation is shown in Fig. 13.2, where the common world price ratio is p_w and where the common production point is Q. Figure 13.2 is drawn so that trade is balanced, or so that the triangles C_hB_hQ and QB_fC_f are identical.

From the analysis of Fig. 13.2 we can conclude that Country H, the country whose tastes are biased toward commodity Y, will import commodity Y and export commodity X, while Country F, where tastes are biased toward commodity X, will import X and export Y. Thus, we conclude that when trade is caused by taste differences, the country will import the commodity toward which its tastes are biased. Another way of stating this result is simply that nations tend to import the goods that are most preferred in consumption when differences in tastes are the predominant source of comparative advantage.

In the specific example given in Fig. 13.2, both countries enjoy gains from trade. The gains-from-trade theorem does apply to this particular

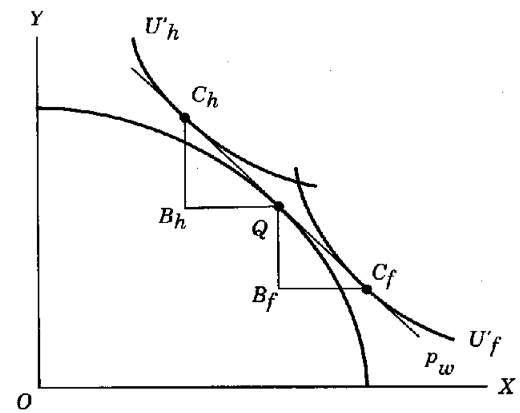


FIGURE 13.2
Trade based on differences in tastes.

model because none of the assumptions required for the theorem is violated. Similarly, we can see that factor prices will be equalized by trade, for at the international trade equilibrium, both countries produce at exactly the same point. Because they share the same production functions, the two countries will have precisely the same efficiency locus, and the equilibrium production same point on this locus. As a result, relative and absolute factor prices must be the same in the two countries. The Stolper-Samuelson theorem also holds in this model. In Country F, for example, trade results in a relative reduction in the price of X. If X is labor-intensive, the wage-rental ratio will fall, the real return to labor will fall, and the real return to capital will rise. The opposite changes will occur in Country H. Thus, the primary results from Chapter 8 apply to a model in which trade is due to taste differences.

Nonhomogeneous Tastes

An alternative possibility, which also generates differences in international consumption patterns, is that tastes are nonhomogeneous but are nevertheless identical in both nations. This nonhomogeneity will help explain the fact that budget shares for food and other consumption items vary widely across countries at different levels of income per capita.

Recall that in Chapter 3 we discussed a situation in which, although tastes were nonhomogeneous, aggregation of the individual curves into community indifference curves was possible. This occurred when income-consumption curves were linear but did not go through the origin. We use this type of "quasi-homogeneous" preferences in Fig. 13.3 to illustrate how differences in per capita income lead to differences in consumption patterns, which in turn lead to trade. Assume that we have two countries with identical populations but that Country F has uniformly superior technologies for producing both goods X and Y. In Fig. 13.3 the production possibility curve of Country F ($T_f T'_f$) is a "radial blowup" of Country H's production

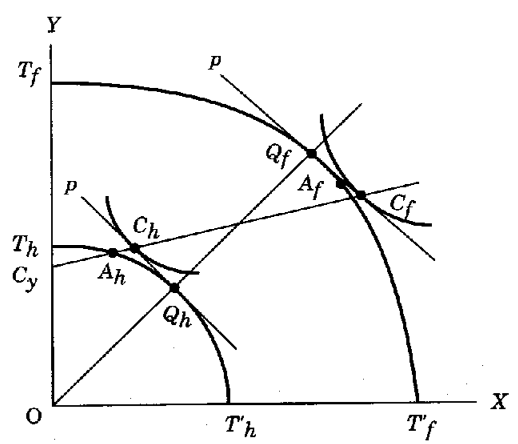


FIGURE 13.3
Trade based on nonhomogeneous tastes.

possibility curve ($T_h T'_h$). That is, along any ray from the origin, the slopes of the two production possibility curves are the same. Assume also that tastes are nonhomogeneous in that there is a "minimum consumption requirement" of Y of the type discussed in Chapter 3. The origin for a system of indifference curves is then point C_y. All consumers in both countries are assumed to have the same (nonhomogeneous) preferences.

Given that the populations are identical by assumption, the larger national product associated with the higher production possibility frontier implies that Country F has higher per capita income than does Country H. The lower income per person in Country H implies a relatively high demand for Y in that country in autarky and vice versa for Country F. The autarky equilibria will be at points like A_h and A_f in Fig. 13.3, where because of the similar production structure, there is a relatively high autarky price of Y in Country H and a relatively high autarky price of X in Country F. As in the previous section, each country will have a comparatively high price for its "most preferred good," except now, "preference" derives from differences in per capita income. Underlying preferences or tastes are assumed to be the same for all consumers.

The autarky price differences lead to a trading equilibrium as shown in Fig. 13.3. The production points of countries H and F are Q_h and Q_f, respectively. These lie along the same ray from the origin because of our "radial blowup" assumption made earlier and the equalization of commodity prices by trade. Consumption points are C_h and C_f for countries H and F, respectively, at the free trade price ratio p. These consumption points lie on a ray through C_y. Country H, low in per capita income, imports Y, and Country F, high in per capita income, imports X. As in the case of the taste differences shown in Fig. 13.2, trade is due to differences in demand. In both examples, the countries produce goods in the same proportions, but they consume them in different proportions at the same set of prices.

Notice an important implication of this analyis. With quasi-homogeneous preferences of this kind, the proportion of income spent on good Y (the good with the minimum consumption requirement) falls and the proportion spent on good X rises as per capita income increases. Another way of stating this is that demand for Y is income-inelastic (elasticity is less than unity) and demand for X is income-elastic (elasticity is greater than unity). Comparing this to the trade pattern established previously, we see that the poorer country imports the good with the low income elasticity. Strictly in terms of demand patterns, then, we would expect that poorer countries would tend to import food to satisfy their relatively greater demand for it.

The models in this section have demonstrated that variations in tastes, embedded in a standard trade model, cause nations to import the goods they most prefer. Trade in this view is still inter-industry in nature, as it was in the Heckscher-Ohlin model. However, this outcome is at odds with two observations about trade in the real world. First, some countries tend also to export goods for which there is a sizeable domestic demand. Second, international trade in manufactures often is more intra-industry in

nature. We can use other ideas about demand patterns to build toward an understanding of these facts.

13.3 THE LINDER HYPOTHESIS

This analysis of the possible role of per capita income in determining trade leads naturally to a discussion of the ideas of Swedish economist Staffan Linder (1961). Linder argued that the principles governing trade in manufacturing goods are not the same as those governing trade in primary products. He was quite prepared to support the idea that trade in primary products is determined by factor endowments. However, he argued against the notion that differences in factor endowments are the major determinants of trade in manufactured products. He chose instead to highlight the role of demand, beginning his argument with an observation similar to the one we made in our chapter on increasing returns to scale: a large volume of trade exists between the developed countries. These countries have very similar factor endowments and thus, according to the Heckscher-Ohlin theory, we might not expect a large volume of trade between them. We must, therefore, look for a cause of trade other than factor endowments.

Linder contended that a manufactured good is created by an innovative entrepreneur in response to a perceived demand. A new manufactured good is, in other words, introduced only when an entrepreneur believes there is sufficient potential demand to warrant production. It is this perception of potential demand, rather than considerations of factor endowments, that triggers production. The second assertion in the argument is that an entrepreneur is most familiar with the home market. Barriers of distance, language, and culture mean that entrepreneurs are much less likely to be able to perceive what kinds of new products could be successfully introduced into foreign markets.

The third assertion follows from the first two. For a manufactured product to be produced in (and therefore potentially exported from) a country, there must exist significant home demand for the product. The range of manufactured goods produced in a country is therefore determined by domestic demand as much as by production considerations such as factor endowments.

Suppose that an entrepreneur perceives a home demand for a product and begins production. Where will this entrepreneur find export opportunities for additional sales? Linder argued that the best opportunities will be found in countries that have very similar demand patterns to the entrepreneur's home country. Thus, if an American invents a new communications device, such as a modem or telefax machine, and produces it for the home market, the best export opportunities for this device will be found in Canada, Western Europe, and Japan. Similarly, entrepreneurs in Canada, Western Europe, and Japan will find that their best foreign markets will be in the United States and in one another's countries.

The final assertion in Linder's argument is implicitly contained in this last statement. The countries with the most similar demand patterns for manufactured goods will tend to be those with similar per capita incomes. People in countries with lower per capita incomes may wish to buy relatively simple products. However, people in countries with much higher per capita incomes may want more sophisticated devices, such as coffee makers with flashing lights, digital readout, and remote control. Thus, the volume of trade in manufactured goods will be highest among countries of similar per capita incomes, such as the United States, Canada, Western Europe, and Japan. For obvious reasons, Linder's hypothesis is also referred to as the *overlapping demand hypothesis* or the *preference similarity hypothesis*. Predictions about which specific products each country might export are difficult to make, because such exports depend on the history of entrepreneurial activity in each market. Overall, however, international patterns of income and demand determine the extent of trade in manufactured goods. In particular, it is possible to export goods for which there is a strong domestic demand.

13.4 EXPLAINING THE EXISTENCE OF INTER-INDUSTRY TRADE AND INTRA-INDUSTRY TRADE: COMBINING DEMAND AND SUPPLY INFLUENCES

So far, we have made a number of references in the text to the concepts of *inter-industry trade* and *intra-industry trade*. We begin this section by discussing these concepts more fully before going on to consider a model in which important determinants of these kinds of trade can be combined into a unified general equilibrium trade theory.

As we have suggested, *inter-industry trade* refers to international exchange of widely dissimilar goods, such as the exportation of automobiles in return for imported clothing. It should be evident that such trade stems from differences in countries that produce rankings of sectoral comparative advantage. In the Ricardian model, for instance, international differences in labor productivity (associated with different technologies) in distinctive types of goods generate a particular trade pattern. More fundamentally, however, economists associate inter-industry trade with the Heckscher-Ohlin model, in which variations in factor endowments by country and in factor intensities by commodity determine comparative advantage. Thus, labor-abundant countries tend to export labor-intensive goods, such as clothing, and capital-abundant countries tend to export capital-intensive goods, such as automobiles. Land endowments are also quite important in determining trade. As noted previously, it follows that inter-industry trade should be prevalent in trade between countries with disparate endowments, such as labor-abundant, developing countries on the one hand and capital-abundant, developed countries on the other. Except in cases of peculiar

endowment structures, such as the acute scarcity of natural resources and land in Japan, we would not expect to see much trade among the developed economies as a group.

It should be clear that this last prediction is false. As we discuss in Chapter 14, trade among developed nations accounted for 57.0 percent of world exports in 1990. Most of that trade was in manufactures involving *intra-industry trade* (IIT), or the simultaneous importing and exporting of similar products. There is substantial IIT among the industrialized nations both in manufactured inputs, such as machinery, high-technology electronics, and specialized chemicals, and in manufactured final goods, such as automobiles, consumer durables, cosmetics, and alcoholic beverages. The United States, for example, is a significant importer and exporter of both wines and beers. It is difficult to reconcile this fact with theories of trade based on differences in factor endowments.

Explanations for Intra-Industry Trade

There are numerous possible sources of IIT. We limit ourselves here to four explanations that are seen as the most important.[4]

Industrial classification. Some trade flows are misleadingly measured as IIT because of the industrial classification system. For example, the United States exports fruits and vegetables during the summer, when they ripen, and imports produce in the winter from countries in the southern hemisphere. This trade, based on *seasonal growing variations,* could easily be explained in a Heckscher-Ohlin framework by considering climate as a factor of production. However, trade flows are typically reported on a calendar-year basis, so both imports and exports of identical goods exist within these trade categories. A second and more significant example lies in the common phenomenon of high-wage nations' producing sophisticated components of electronics products, exporting them to low-wage nations for assembly into final goods, then importing the final products. A prominent case of this *mid-product processing* occurs in United States trade with Mexico, wherein several American firms ship components to *maquiladora* plants in Mexico along the border for assembly into automobile parts, televisions, and the like.[5] Similar processes happen in trade between the high-income countries of the European Community and the Mediterranean nations and between Japan and developing nations in Asia. The point here is that such trade is largely based on differences in wage costs, suggesting that it is generated by Ricardian or Heckscher-Ohlin factors. Nonetheless, both the exports of components and the imports of final goods tend to be counted in the same trade categories, resulting in IIT.

Transport costs. For many products, shipping costs are high in relation to their market value. The prototypical examples are heavy and inexpensive goods, such as cement, bricks, and lumber, though we could also mention electricity and certain tradeable services like construction. High transport costs imply that markets for such goods are limited geographically. It is quite possible that localized markets exist across national borders, generating two-way trade. For example, it may well be that lumber is exported from British Columbia in Canada to the state of Washington in the United States and simultaneously exported from Maine in the United States to Quebec in Canada. There are numerous examples in Western Europe of cities that lie across a national border from each other, meaning that these cities form an integrated market. Two-way trade associated with transport costs is common as different localized binational markets engage in ordinary commerce.[6] In this sense, IIT can be a result of political and geographic decisions on where to draw national borders.

Product differentiation. One of the strongest assumptions in traditional trade theories is that imported and domestic goods are completely homogeneous, or perfect substitutes in consumption. However, many commodities, especially manufactured goods, are differentiated by style and quality. It is obvious that automobile models vary in size, horsepower, comfort, performance, and appearance. Similarly, wines, beer, furniture, and other goods are differentiated by quality and variety. Manufactured inputs, such as machinery, electronic components, and software are also differentiated in significant ways. If preferences are heterogeneous, so that imported goods and similar domestic goods are seen by consumers as imperfect substitutes, the products can command different prices and profit markups in various markets. This is because each variety of a product will face a distinct demand curve in each market, allowing its producer to act with some market power. Thus, firms will compete not only on the basis of price, but also on the bases of performance, quality, service, and other dimensions of product differentiation.

As we discussed in Chapter 12, these simple observations carry crucial implications for international trade. It is evident that consumers and firms prefer to have many choices of which car or wine or machine to purchase. When either the *love of variety* idea or the *ideal product* idea is incorporated into models of monopolistic competition and product differentiation, it follows that international trade is simply an extension across borders of tastes for diversity. Thus, there are consumers in the United States who prefer German beer and consumers in Germany who prefer American beer, resulting in two-way trade. Similarly, there are homeowners in Canada who enjoy Scandinavian furniture and homeowners in Sweden who prefer Canadian furniture. Trade in differentiated inputs is also prevalent. Thus, tastes for variety and the ability to differentiate products provide a powerful explanation for IIT.

Increasing returns to scale. IIT associated with product differentiation is considerably reinforced by the existence of scale economies. We developed numerous perspectives on scale economies and trade in Chapter 12. For our present purposes, we simply reiterate the fact that if industries in each

country tend to produce a relatively small range of differentiated goods, each subject to increasing returns, the result will be substantial two-way trade at lower costs. It is inefficient for each country to produce an entire set of models of automobiles and trucks, for example, if some models can be obtained from other countries. Thus, each automobile industry in the United States, Canada, Germany, France, the United Kingdom, Japan, and Korea produces a subset of differentiated models at relatively low cost. Indeed, in nearly any manufactured product for which scale economies are potentially important, the combination of increasing returns and product differentiation leads to significant IIT.[7] Finally, note that increasing returns clearly expand the likelihood that a nation will export goods for which it has strong domestic demand.

A Unified Model of Trade

Many of the various elements that we have analyzed in this and the preceding chapter can easily be combined into a model that helps explain the stylized facts that motivated our discussion: the apparently large volume of trade among the industrialized countries relative to trade between the industrialized countries and the developing countries, and the existence of IIT in manufactured goods.[8]

Suppose that there are two goods, food and manufactures, and that food is labor-intensive while manufactures are capital-intensive. Assume that food has a high minimum-consumption requirement (implying a low income elasticity of demand), whereas manufactures have no minimum-consumption requirement (implying a high income elasticity of demand). Let there be two blocks of countries, a capital-abundant "North" and a labor-abundant "South." (These titles, which are used commonly in international economics, are meant to distinguish the industrialized, higher-income nations from the developing, lower-income nations.) The North will be relatively specialized in producing manufactures because of its capital abundance, as suggested by the factor-proportions theory. Suppose also that the countries of the North have higher per capita incomes, which is the result when ownership of capital is spread uniformly over laborers. Countries of the North will therefore consume relatively more manufactures because of their higher incomes. In contrast, the labor-abundant South will be relatively specialized in both producing food, because it is labor-intensive, and consuming food, because of the lower per capita incomes. If there were no taste biases, the South would export food and the North would export manufactures exactly as predicted by the Heckscher-Ohlin theorem. However, the consumption biases mean that the level of North-South trade in food for manufactures will be reduced below what it would be in a Heckscher-Ohlin world. The consumption bias in each region toward its own export good reduces both imports and exports.

Assume finally that manufactures are not a homogeneous good but a collection of differentiated manufactured goods, as discussed earlier. With increasing returns in each product, each manufacturing firm will produce a somewhat unique good and sell it to consumers in all the countries that make up the industrial North. This is intra-industry trade because these countries are simultaneously importing and exporting similar manufactured goods to one another. In the North, the taste bias toward manufactures due to the high per capita income now actually increases the amount of intra-industry trade among the northern countries. Relative to a Heckscher-Ohlin world with homogeneous tastes, the northern manufacturing firms switch exports from the South, where demand is low due to low per capita incomes, to other northern countries where demand is conversely high. Thus, nonhomogeneous demand leads to a decrease in North-South trade and to an increase in IIT among the northern industrialized countries. These are the stylized facts that were to be explained. We should note that while this model is most useful from the standpoint of economic realism, it makes the assumption that there is a strong positive correlation between capital intensity in production and a high income elasticity of demand in consumption. This assumption, which seems reasonable on some grounds, has not yet been supported by conclusive, empirical evidence.

13.5 THEORIES OF TRADE BASED ON DYNAMIC CYCLES

We turn next to issues of dynamic changes in trade patterns. During the 1960s and perhaps earlier, some international trade economists observed that the pattern of trade in the world had undergone a gradual shift during the 20th century. The change seemed to be particularly pronounced with respect to the manufactured exports of some developed countries. It was noted that while numerous goods were originally developed and produced in the United States, the location of their production often shifted subsequently to Europe and then to developing countries, only to be replaced by a new generation of products in the United States. This gradual and systematic pattern of change seemed to call for a new theory of international trade that could account for the dynamic introduction of new goods and the movement over time in comparative advantage toward the developing countries as the new goods advanced in age.

The Life Cycle for New Products

Several economists developed informal theories of this process as related to new products, though they attempted to relate it to notions of underlying comparative advantage.[9] Raymond Vernon was the first to construct an explicit *product cycle* hypothesis.[10] In this view, a product goes through a full life cycle from innovation to standardization. Vernon's conception of the *innovative stage* of the cycle borrowed ideas from the Linder hypothesis, though there is nothing particularly dynamic (having to do with changes over time) in Linder's analysis. Consider a product such as a personal computer (PC). At some point, entrepreneurs in some high-income country, such as the United States, Germany, or Japan, decide that potential demand

has risen to the point where PCs can be profitably introduced. High incomes are again important to the extent that new products are income-elastic in demand. It is also important for demand in the innovating country to be fairly price-inelastic for the new good, in order for the firm to be able to earn some (at least temporary) profits as it refines the production technique. Thus, demand patterns are again central to the theory.

Production of the computers initially occurs in the home market where there is easy communication between factory managers and sales personnel. This localized production, according to Linder and Vernon, is necessary in order to facilitate critical revisions of the PC and of the technology for producing it. Such revisions to the innovative product can benefit from customer comments, while providing repair services is easier in local markets. Exports to other high-income countries are expected to begin within a relatively short time as consumers in those countries recognize the existence of the PC. An increasing proportion of domestic output will likely be exported to other countries as their incomes grow.

The second phase, often called the *maturing stage* of a product cycle, occurs when the technology for producing the good has become fairly stable, and demand in other high-income and middle-income countries has risen to the point where entrepreneurs there find it profitable to begin production themselves. After all, they have the advantages of lower transport costs and local communication, while the limited need for additional technical change in the personal computer has diminished. Such production could come about through several mechanisms. First, the firm that innovated the PC could undertake foreign direct investment in factories in Western Europe and countries such as Canada, Australia, and even South Korea or Taiwan. Indeed, the product cycle model provides one hypothesis about why firms engage in foreign direct investment. Second, local firms in these countries could begin production of the new good under licensing arrangements with the innovating firm. Finally, firms in these other countries might develop a competing product with similar characteristics, such as a PC with more memory or improved graphics capabilities. After some time it might well be that these firms begin exporting the PC and its substitutes back to the original innovating country. One reason is that they might have lower costs of production because of the lower wage rates in their economies, an explanation of particular relevance to, say, Taiwan. These producers might also gain scale economies and production expertise as they penetrate markets in the innovating nation. Also important is that firms in the innovating country will have moved on to develop newer versions of the good and will have begun scaling back production of the older PC model.

The third phase of the product cycle, often called the *standardized stage,* occurs when production of the original PC becomes routine and labor-intensive. Note that by this stage, demand for the product is likely to have become much less price-inelastic because consumers have come to understand how the product works and numerous competing products have been introduced. In this final phase, production is likely to move to the low-wage developing countries through foreign direct investment, as

there is relatively little local ability in those countries to develop competing products. Such production is capable of servicing markets in North America, Japan, and Western Europe so that the original producers are now in the position of import competitors. The process could even continue to the point where the original producers are driven out of the market entirely.

However, disappearance of the innovating firms is unlikely. As the last phase of the product cycle approaches, our example PC firm has been developing and exporting newer products such as work stations or supercomputers. There is a continuous introduction of new products, each of which may be expected to go through a similar life cycle. As a second example, color television receivers were first developed decades ago by American firms. However, Japanese and European televisions began soon thereafter to compete with North American firms and eventually dominated the North American market. Currently, most television sets sold in the United States are produced in developing countries, such as Mexico, under license to U.S., Japanese, and European firms. However, North American electronics firms have moved on to other products and are competing with firms in Japan to be the first to introduce commercially successful high-definition television receivers.

Another example might be textile production. During the early nineteenth century, Great Britain had a large share of the world export market in high-quality woven textiles. As the century progressed, many of Britain's export markets were lost to new producers in North America and Europe. By the middle of the twentieth century, these countries were in turn losing sales to producers in Japan. The Japanese began to experience cost increases in the 1960s, and production began to shift to countries such as Korea, Taiwan, and Mexico. By this time the Europeans, North Americans, and Japanese were beginning to specialize in new products, such as sophisticated textile materials, that may someday also be produced largely by developing countries.

The Life Cycle for New Technologies

It seems useful to distinguish between new products, on the one hand, and new technologies as embodied in production processes (such as steelmaking) or new producers' goods (such as machinery) on the other hand. The product cycle previously described tends to focus on final consumption goods. Yet it seems likely that there is a closely related phenomenon with respect to production technology, which we might call the *technology cycle.* Techniques of production and various types of machinery often seem to follow a cycle from development and use in the advanced countries to eventual use in developing countries. Such a phenomenon is of interest because production technologies and producers' goods are important exports of the industrialized nations.

A simple theory of a technology cycle begins with the observation that the industrialized nations are high-income and high-wage countries.

High wages create a strong incentive to invest in labor-saving production technology, because the payoff to a new innovation is the number of worker hours saved, times the wage rate. The industrialized nations are thus the leaders in the development of new technology. This new technology improves productivity and tends to lead to further wage increases. Indeed, it would be very difficult to decide whether the development of new technology stimulates wage increases or vice versa. Presumably the causality runs both ways.

In any case, beginning with the observation of continuous technological development and wage increases, we can postulate a technology cycle similar to the product cycle. In the first phase, a new technology, such as a piece of textile machinery, is developed in an industrialized country. This machine is used for production in the innovating country or in other high-wage developed nations. It is not exported to developing countries because its use would make production there inappropriately capital-intensive. As time passes, wages rise in the developed countries to the point where the textile machine no longer permits profitable production. Simultaneously, incomes in some developing nations rise to the point where use of the machine is justified for foreign production. The machine becomes an export of the industrialized country. In later stages of the cycle the machine may, in fact, be produced abroad.

To the best of our knowledge, this phenomenon is not well-researched and should thus be regarded as somewhat speculative at this point. Yet it does seem to offer a potentially important hypothesis about the direction of trade in technology and production equipment. As noted, however, a satisfactory theory must deal with the simultaneous causality between technical change and high per capita income.

Some Implications of Cycle Models

It is important to understand the implications of this continuous process of product and technology development and production relocation. There are basically two forces in operation here: the innovation of new goods and technologies and the international diffusion of information about how to produce them. As demonstrated in theoretical analyses by Paul Krugman and David Dollar, these forces set up a potential conflict between the interests of the advanced, innovative countries (the "North" in our earlier terminology) and those of the developing nations (the "South").[11] The introduction of new goods helps satisfy the tastes for variety that consumers in all countries share in these models. However, production of these new goods is at first feasible only in the North because only firms in that region have the technical knowledge required. This gives Northern firms a temporary monopoly, which is translated into higher wages for northern workers. Thus, new products raise welfare for northern workers through higher wages and more consumption variety. The faster the rate of innovation, the greater the gain in utility for the North. In these models, new products also raise welfare for southern workers because the variety gains outweigh the higher costs of

new goods associated with the premium wages. In general, of course, this may not be true, and the monopoly costs could be larger.

The existence of lower wages in the South sets up an incentive to transfer technological information about production to the South, where production costs would be reduced. However, when technology is diffused in this fashion, the Northern monopoly disappears. Thus, technology transfer raises the welfare of southern workers through higher wages but lowers the welfare of northern workers by reducing their wages (though prices also decline). If rates of technology transfer are quite rapid in comparison with rates of new product development, relative wages may converge. In this sense, Northern countries may have an incentive to enact barriers, such as patents, to free transfer of technology.

While these models are highly stylized, they do point out important features of the world economy. Advanced countries like the United States, Canada, and Japan have high wages due in part to their highly innovative economies. An important component of maintaining high incomes is the need for continuous product and technology innovation through research and development expenditures, assisted by the maintenance of a technically skilled labor force. At the same time, the transfer of new technologies is crucial for raising incomes in the developing countries. Indeed, some countries may find themselves rapidly climbing a "technology ladder" from the production of standardized goods to product innovation. Japan has passed through this process, and Korea, Taiwan, Singapore, and Hong Kong appear to be doing the same now. Other countries, such as Thailand and Indonesia, quickly take their places at the lower end of the ladder. These dynamic processes are a key determinant of global economic growth, but they do put a burden on the wealthy countries to engage in continuous innovation. Failing this, the developed countries could be tempted to erect barriers to imports from the poorer countries, an observation that seems to be increasingly consistent with actual trade policies.

13.6 CYCLE MODELS AND COMPARATIVE ADVANTAGE

At first glance, these dynamic theories seem to stand in stark contrast to our earlier discussions about the determinants of trade. Although the earlier discussions suggested a stable pattern of trade based on patterns of comparative advantage, returns to scale, and so on, the dynamic theories seem to suggest a continuously shifting pattern of trade, at least for manufactured goods. Are we forced to choose between two irreconcilable approaches?

Much of the apparent conflict disappears if we think of goods as being composed of a number of characteristics. For new goods, the standard Heckscher-Ohlin effects of labor or capital intensity are relatively unimportant in comparison with issues such as engineering inputs, shifting technical specifications, and experimentation in use. As products mature, however, we usually observe a relatively stable pattern of trade emerging in product characteristics associated with more traditional factors. Indeed, whether a mature product is produced with stable technologies that

are labor-intensive (say, television production) or capital-intensive (say, steel plate production) determines the type of countries that will export back to the innovating nations.

Similarly, even though the particular goods being produced in a country may continually change, the underlying characteristic of goods exported from a particular country will remain largely the same, as long as that country remains in a similar position relative to others in terms of its wages and its capacity for product and technology innovation. The product-cycle theory suggests that we should find the wealthiest nations specializing in new consumer goods that cater to high-income tastes. These nations should be observed to produce and export new capital-intensive, or labor-saving, equipment and technologies. If the types of product characteristics embodied in a nation's exports are in turn systematically related to the determinants of trade discussed earlier, including factor endowments, scale economies, and nonhomogeneous tastes, then the product-cycle approach may not, in fact, be that different after all.

Several reservations can be expressed about the cycle theories. As we have just mentioned, it may be that the cycle theories are actually not very distinctive after all, if we think of countries as trading the characteristics embodied in goods rather than the goods themselves. Instead of viewing the United States as having moved from exporting radios to televisions to home computers, perhaps we should think of the United States as exporting sophisticated electronic goods, "sophisticated" being defined relative to the period in question.

Second, although incentives for product innovation and imitation remain as strong as ever, changes in the world economy may have shrunk the *length* of typical product and technology cycles so much that the model has limited empirical validity. This is largely because the product cycle theory as developed in the 1960s significantly underestimated the ability of multinational enterprises to move production abroad at little cost. As we noted earlier, the theory relies on the assumption that a new product must initially be produced at home in order to allow easy communication among engineers, plant managers, and sales personnel. In an era of increasingly sophisticated multinational firms, this communication problem has become less of a constraint. Firms may quickly transfer the production of a new product to the location that offers the most attractive factor prices. But in this case, production is determined by the traditional determinants of comparative advantage. We should also point out that in recent years, changes in global production techniques have tended to deemphasize traditional mass-production techniques in many high-technology sectors in favor of more flexible manufacturing methods pioneered by Japanese firms. Where such techniques are prevalent, we would expect to see much less shifting of production and comparative advantage to low-wage countries than we saw in the 1970s.

A final reservation about the cycle theories is that they are largely partial-equilibrium explanations of trade and are stated in such a way that it is difficult to understand the nature of causality. They seem to assert

that high per capita incomes cause the development of new products and technologies. But, as we noted previously, it would be equally valid to assert that new technologies cause high incomes. Until this type of problem can be fully worked out, the cycle models will have to be regarded as incomplete. A satisfactory model should give a complete or general equilibrium description of the economy, rather than treating the level of income as exogenous. In the last ten years, economists have begun to develop such models, in which growth in income levels and incentives for product innovation are endogenously determined as a result of variables such as initial factor endowments, tax and tariff rates, and economies of scale.[12] Interestingly, in these so-called "new growth models," international trade is as much a spur to further innovation as it is the result of underlying demand and supply factors in static models.

13.7 CONCLUDING REMARKS

In this chapter we have examined several ways in which demand influences international trade, and we have considered simple theories of dynamic technical change and trade. Our main conclusions may be stated as follows.

1. Demand differs across countries both because people have different tastes and because demands depend on per capita income when tastes are identical but nonhomogeneous. Empirical studies of demand tend to strongly reject the hypothesis of homogeneity, as we will discuss again in the next chapter. In these relatively simple specifications of demand patterns, countries tend to import the goods for which they have the strongest preferences.
2. Differences in demand caused by differences in per capita income are a cornerstone of the Linder hypothesis. This model has other elements, such as the role of entrepreneurs in developing new products, but per capita income is central to the implications of the models for international trade questions. A country can export only those products for which an entrepreneur has perceived and filled a domestic demand. Thus, trade in manufactures is most prevalent among developed countries, which have similar tastes because of similar per capita incomes.
3. The existence of some intra-industry trade is due to the categorization of industries for purposes of reporting trade data. Some of it is due to transport costs in conjunction with localized markets that spill over national borders. Most importantly, however, IIT is the result of preferences for product variety in conjunction with economies of scale.
4. These factors may be combined with the notion of nonhomogeneous preferences and factor endowments into a unified model that explains the prevalence of trade in differentiated manufactures among developed countries and the prevalence of inter-industry trade between developed countries and developing countries.
5. The product-cycle model also depends heavily on differences in demand associated with per capita income. Fundamentally, the cycle models

are useful in understanding the importance of product and technology innovation and international diffusion. These processes have important influences on incomes in different nations.

6. The cycle models may be criticized on numerous grounds. A more satisfactory dynamic theory of trade awaits further development of models that capture the mutual dependence of technical change and per capita income.

PROBLEMS

1. Suppose in Fig. 13.2 that over time, tastes in Country H shift toward commodity *X*. How will this affect welfare in the two countries? How will relative and real factor rewards be affected?

2. In Fig. 13.2, how is the volume of trade affected by the degree of differences in tastes between the two countries?

3. In Fig. 13.3, show that trade disappears as the minimum consumption requirement for *Y* goes to zero.

4. Try to think of several products that have undergone cycles. What are the underlying characteristics of the products in question?

5. Using the Linder hypothesis as applied to the product cycle, explain why the United States was a leader in the development and export of automobiles. Explain why high-quality hockey skates were first produced in Canada.

6. In which ways is the product-cycle model inconsistent with the Heckscher-Ohlin model? In which ways are the two consistent?

7. How would you use the product-cycle model to help explain the existence of IIT?

NOTES

1. Our review in the next chapter of empirical work in trade considers one such study, the results of which suggest that demand differences do indeed play an important role.
2. The source for these data is World Bank, *World Development Report*, 1993.
3. Deeper analysis of trade and economic growth will be presented in Part Four.
4. Interested readers are referred to Greenaway and Milner (1986). We also provide further commentary on the empirical importance of IIT in the next chapter.
5. In fact, this process is encouraged by certain provisions of U.S. and Mexican trade policy. Under U.S. law, only the value added by workers in Mexico is subject to import taxes when the final goods are shipped back. These policies will be phased out under the terms of the North American Free Trade Agreement.
6. Readers should not infer from this that increases in transport costs would expand IIT. Rather, higher transport costs tend to diminish all trade flows. However, we do observe two-way trade in particular products with significant shipping costs.
7. It is also possible to develop models of IIT in homogeneous products manufactured under increasing returns. Recall from Chapter 12, for example, that if two identical firms produce the same good monopolistically in autarky in two countries, the introduction of free trade would lower prices in a Cournot-Nash equilibrium through the *threat* of two-way trade or cross-penetration of each other's markets. However, in the equilibrium, there could be IIT if there were no transport costs. Such "cross-hauling" is possible even in the presence of transport costs if the firms engage in "reciprocal dumping" or charging lower prices in export markets than at home, as shown by Brander and Krugman (1983).
8. This model is based on the formal theory developed by Markusen (1986). Important elements of this theory were also advanced by Krugman (1980).

9. See Posner (1961) and Hirsch (1967).
10. See Vernon (1966).
11. See Krugman (1979) and Dollar (1986).
12. Grossman and Helpman (1991) provide a series of such models and a review of the literature.

REFERENCES

Brander, J. A., and Krugman, P. R. (1983). "A 'Reciprocal Dumping' Model of International Trade." *Journal of International Economics* 15: 313–323.

Dollar, D. (1986). "Technological Innovation, Capital Mobility, and the Product Cycle in North-South Trade." *American Economic Review* 76: 177–190.

Greenaway, D., and Milner, C. (1986). *The Economics of Intra-Industry Trade*. London: Basil Blackwell.

Grossman, G. M., and Helpman, E. (1991). *Innovation and Growth in the Global Economy*. Cambridge: MIT Press.

Helpman, E., and Krugman, P. R. (1985). *Market Structure and Foreign Trade*. Cambridge: MIT Press.

Hirsch, S. (1967). *Location of Industry and International Competitiveness*. Oxford: Clarendon Press.

Linder, S. B. (1961). *An Essay on Trade and Transformation*. Stockholm: Almqvist & Wiksell.

Krugman, P. R. (1979). "A Model of Innovation, Technology Transfer, and the World Distribution of Income." *Journal of Political Economy* 87: 253–265.

———. (1980). "Scale Economics, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70: 950–959.

———. (1981). "Intra-Industry Specialization and the Gains from Trade." *Journal of Political Economy* 89: 959–973.

Markusen, J. R. (1986). "Explaining the Volume of Trade: An Eclectic Approach." *American Economic Review* 76: 1002–1011.

Posner, M. V. (1961). "International Trade and Technical Change." *Oxford Economic Papers* 13: 323–341.

Vernon, R. (1966). "International Investment and International Trade in the Product Cycle." *Quarterly Journal of Economics* 80: 190–207.

———, ed. (1970). *The Technology Factor in International Trade*. New York: Columbia University Press.

CHAPTER 14

EMPIRICAL STUDIES OF COMPARATIVE ADVANTAGE MODELS

14.1 INTRODUCTION

The preceding chapters demonstrate that economists have developed numerous models to explain why international trade takes place. These models make assumptions that, in large measure, embody strong abstractions from reality in order to isolate the particular influence of some important variable on the pattern and volume of trade. These theoretical abstractions allow analysts to investigate the implications of different circumstances for trade flows and economic welfare. Undertaking an analysis in the real-world environment of imperfect competition, numerous factors, commodities, and trade restrictions would be most difficult. Nonetheless, it is natural for analysts to wonder how well their theoretical predictions correlate with actual empirical data on international trade. Accordingly, a large body of literature has appeared in which economists attempt to test various aspects of the theory of comparative advantage or to assess the importance of different explanations for trade. In this chapter we review only a small subset of the most important studies.[1]

We first discuss some complications that should be kept in mind in considering this body of empirical work. One problem is that it is difficult, even in principle, to test theories of comparative advantage directly because they rely on statements about differences in autarky relative costs and

prices across countries. Autarky, or the absence of any trade, is virtually an unobservable situation, and available data would be influenced by international trade. Thus, economists resorted to indirect means of testing trade theories based on observable variables.

A second problem is that to establish theoretical statements about trade, economists make numerous simplifying assumptions that cannot be true under all realistic circumstances. Thus, even in cases where it is possible to translate our theories into equations that embody observable variables, these equations cannot be expected to hold literally or without error. Accordingly, because they cannot fully test trade theories, empirical trade analysts must pose a simpler question: "How closely do actual trade data correspond to the levels predicted by various trade theories?" Given this constraint, empirical work consists largely of measurement and judgment rather than precise testing.[2]

A third point is that our various international trade theories should not be seen as competing hypotheses. Rather, each theory tends to focus on a particular aspect of national economies that is expected to induce trade. In principle, each of these influences operates simultaneously, both alone and in conjunction with the others, to explain the pattern and volume of trade. The task for empirical economists, therefore, is to assess the relative importance of various trade determinants.

Despite these problems, economists have made great progress in studying the effects of various influences on the patterns of international trade. It is well worth examining some of the important work on the determinants of comparative advantage. One reason for considering this work is that it has caused some trade theorists to reconsider their models in light of new empirical regularities discovered in trade data. A second reason is that, whatever their methodological shortcomings, our empirical studies have resulted in a large body of evidence that is highly suggestive of the importance of varying determinants of trade. Finally, as our theories have suggested, understanding the sources of international trade is critically important in evaluating the relationships between trade and such significant issues as income distribution, market structure, and technology. In turn, empirical work can be used to inform the decisions of governments making trade policy.

14.2 THE GAINS FROM TRADE

With few exceptions, our models point strongly to the conclusion that a country in a state of free trade likely enjoys substantially higher welfare than it would in isolation. This view was perhaps most apparent in the Classical and Heckscher-Ohlin models, in which the ability to trade at world prices allowed nations to specialize their resources along the lines of comparative advantage and allowed consumers to avail themselves of cheaper import prices. Additional gains from trade came from rationalization of industry, greater economies of scale, improved quality, and enhanced competition.

As mentioned earlier, it is difficult to test for the existence of gains from trade because of the absence of autarky data. Economists often appeal to an obvious point in claiming that there are such gains: countries could presumably refuse to trade if engaging in trade made them worse off. Thus, from the standpoint of *revealed preference,* countries consider themselves to be better off in a regime of open trade than in isolation, which was the basic point of Chapter 5. Of course, we must recall that free trade may well redistribute income, making some individuals worse off and some better off, so that some scheme of compensatory payments to the losers must be potentially available.

In practical terms, it is not difficult to see the wisdom in the view that free trade is superior to autarky, because a policy of self-sufficiency is generally no more sensible for a country than it is for an individual. If a small economy, such as that in Hong Kong or Switzerland, attempted to meet all its needs for food, machinery, steel, and the like through domestic production alone, success would come only at high costs, tending to impoverish the nation.

This powerful statement is supported by historical evidence in a fascinating study by Richard Huber.[3] Japan moved from a position of virtual autarky in 1858 (the end of its feudal period) to one of nearly free trade in the 1870s after the Meiji Restoration. The effects on the Japanese economy were dramatic. In this period, foreign trade rose from a negligible percentage of national income to perhaps 7 percent. In autarky Japan needed to produce both primary goods and manufactures. Japan was especially inefficient in the production of the latter; in order to sustain production of iron bars, for example, the price of that commodity relative to tea was some 9 times the corresponding relative price in London. Once trade began, the prices of tea and silk, Japan's primary export goods, rose dramatically toward world levels. Similarly, the relative prices of import goods fell sharply as Japanese consumers and firms took advantage of lower foreign costs. Overall, Huber calculated that by the 1870s, Japan's terms of trade had improved by 340 percent. This change, in conjunction with access to better foreign technologies, expanded Japanese real national income by as much as 65 percent in 15 years.

The Japanese example is unique perhaps only in providing an opportunity to study a transition from autarky to open trade.[4] Economists point to a more recent body of evidence, however, that suggests that developing economies that are more open to trade undergo economic transformation and growth more rapidly than those developing countries that are relatively inward-looking and protectionist. Numerous studies have shown that, since the 1960s, the export-oriented economies of Asia, such as Korea, Hong Kong, Taiwan, and Singapore, have performed markedly better than countries in Latin America and Africa. The key difference is that the former countries have generally exposed their industries to competition at world relative price levels, while the latter nations have sustained relative prices that do not correlate well with comparative advantage. Substantial debate continues

among economists over the conclusiveness of such comparisons.[5] However, there seems little question that the more open economies have raised their levels of per-capita income relative to the other countries. This fact suggests that the existence of pervasive barriers to trade tends to prevent nations from achieving full gains from trade.

14.3 TESTS OF THE RICARDIAN MODEL

Most empirical work has focused on the predictions of particular models rather than on the broader question of the gains from trade. The Ricardian model, for example, rests on the assumption of different technologies in different countries, generating varying labor productivities. These labor productivities determine comparative advantage. Tests of the model attempt to find relationships between relative labor productivity and international trade flows.

Note that the sharpest prediction of the theory, that countries are specialized in the goods they export, may be rejected in practice. With few exceptions, countries also produce the goods they import and a host of nontraded goods.[6] Nonetheless, it is interesting to examine how strongly differences in labor efficiency correlate with exports. The pioneering work was done in the 1950s by G. MacDougall, who computed simple measures of average labor productivity in the United States and the United Kingdom for the year 1937.[7] He hypothesized that, given that the American wage rate at that time was approximately twice that in Britain, U.S. firms should have an export advantage in manufacturing sectors for which U.S. labor productivity exceeded twice the level in the U.K. He tested this notion by calculating the ratios of U.S. exports to U.K. exports of 25 products to countries other than themselves. MacDougall considered only trade with third countries because trade barriers greatly influenced bilateral trade between the United States and the United Kingdom. However, exporters in both countries faced largely equivalent market conditions in other countries and could compete on an equal footing. MacDougall's test results were supportive of the Ricardian model. Twenty of the 25 products satisfied the simple prediction that, in cases where U.S. productivity exceeded twice the U.K. level, the ratio of U.S. exports to U.K. exports exceeded one, while in other cases the ratio was less than unity.

A fuller examination along these lines was performed by Robert Stern, who compared American and British trade in 1950 and 1959.[8] By 1950, average U.S. wages were approximately 3.4 times the average U.K. wages, suggesting that the ratio of American to British exports would be greater than unity in sectors where the ratio of outputs per worker exceeded 3.4 and less than unity in other cases. Of the 39 sectors considered, 33 conformed to this prediction in 1950, with the relationship becoming somewhat weaker by 1959. The results for the 1950 data suggested that a 1-percent increase in the ratio of labor productivities was associated with a 1.27 percent percent rise in the export ratio.

These two studies seemed to provide encouraging support for the Ricardian model. In fact, it is surprising that economists have not devoted much additional effort to such analysis in order to verify more conclusively that labor productivity differences constitute an important determinant of international trade. However, in interpreting these results, we must keep two important caveats in mind. First, the empirical specifications were exceedingly simple and did not control for the potential effects of other determinants of trade, such as transport costs, imperfect competition, and product differentiation. Second, the results found by MacDougall and Stern are consistent with other trade theories as well. For example, it is easy to show that in a world where trade is caused by differences in factor endowments but where factor prices are not equalized, the relative productivity of labor will tend to be higher in capital-abundant countries. Accordingly, the results may have simply captured the effects on trade of American capital abundance and British labor abundance in that period.

14.4 TESTS OF THE HECKSCHER-OHLIN MODEL

Recall that the Heckscher-Ohlin model makes a series of strong assumptions in order to isolate the effects of different relative factor endowments on trade between two countries. These assumptions include identical technologies with constant returns to scale, perfect competition, the absence of factor-intensity reversals, identical and homogeneous preferences, the absence of international factor migration, and the absence of impediments to trade. Again, it is obvious that this set of assumptions does not hold in reality and that the predictions of the model cannot be expected to hold literally. The task is to assess the significance of endowment differences in explaining trade patterns. This complicated subject occupies our discussion in this section.

The Leontief Paradox

As we mentioned in Chapter 8, the Heckscher-Ohlin model is incapable of predicting that any country will export a particular good when there are more goods than factors. For example, the more capital-abundant country may not export the most capital-intensive good when there are three commodities. However, the implicit international trade in factor services must obey the Heckscher-Ohlin theorem. That is, with two factors and two or more goods, the capital-abundant country will find that its bundle of exports is more capital-intensive than its bundle of imports.

This theorem on the factor content of trade was first examined by Wassily Leontief in perhaps the most famous empirical study in economics.[9] Leontief had greatly aided the American planning efforts in World War II by developing a technique of accounting for all the inputs required in the production of GNP. This technique, called input-output analysis, recognizes

that the production of, say, an automobile requires *primary* inputs, such as capital and labor, in addition to *intermediate* inputs, such as steel, paint, glass, and the like. The prior production of these latter inputs also requires capital and labor in addition to other intermediate inputs, which in turn require capital and labor, and so on. Leontief developed a method for assembling these various inputs into an input-output table that could be used to compute the total labor and capital embodied in production of any bundle of goods.

An obvious application was to discover how much capital and labor were required to produce U.S. exports in comparison with U.S. imports. Note that an immediate methodological problem arises. While it is sensible to use an American input-output table to compute the factor contents of U.S. exports, a computation of the factor contents of U.S. imports would require detailed and consistent data on production techniques in all foreign trading partners. This was not feasible for Leontief, so he calculated the capital and labor required to produce U.S. goods that are similar to (or compete with) American imports. This procedure is theoretically valid under either of two conditions. First, if the factor-endowments model holds and international factor prices are equalized, each country shares the same techniques of production, and using the U.S. table will not bias the import computations. Second, if production functions exhibit fixed coefficients, or a constant ratio of capital to labor regardless of the factor-price ratio, use of the U.S. techniques similarly captures foreign production methods adequately. Leontief chose the latter condition.

Leontief calculated the capital and labor requirements in the production of a representative bundle of $1 million worth of both exports and import-competing goods in 1947. In that year the United States was unquestionably the most capital-abundant nation in the world and was certainly capital-abundant and labor-scarce relative to the rest of the world. Thus, the expectation was that exports were capital-intensive. Nevertheless, Leontief discovered that the capital-labor ratio in U.S. imports exceeded that in U.S. exports by some 23 percent. This unexpected outcome has been termed the *Leontief paradox.*

Leontief's famous result caused great surprise among economists schooled in the Heckscher-Ohlin tradition. Leontief himself was puzzled by the finding and asserted that the issue was really one of measurement. In particular, his belief was that, because of superior education and training in conjunction with better management techniques, American labor was perhaps three times more productive than foreign labor. Thus, in effective labor units, the United States was really labor-abundant. While this view anticipated later thinking about labor skills, it was ad hoc and unconvincing of the time. The three-to-one ratio suggested by Leontief was not predicated on a careful evaluation of labor efficiency throughout the world; it was simply the ratio required to get the "expected" result. In fact, there is little evidence that this huge superiority in U.S. labor ever existed, even in

the early 1950s. Moreover, to the extent that labor was enhanced by better American management or entrepreneurship, we would expect the relative productivity of capital to be enhanced as well.

Numerous attempts have been made to verify Leontief's results, with mixed results. Because the United States may have been an unusual country, similar computations have been made for other countries with all manner of endowments, incomes, and market structures. In some cases, the results of these studies were consistent with expectations under the endowment model. For example, Peter Heller demonstrated that Japan's international trade in the 1960s followed an interesting dual structure.[10] Japan's exports to less-developed nations were capital-intensive, and its imports were labor-intensive, while its exports to more-developed nations were relatively labor-intensive in comparison with imports.In many other studies, however, the results seemed to contradict the Heckscher-Ohlin theorem. Geographically, then, the Leontief paradox is not an isolated event.

A more compelling objection was that 1947 was not a very appropriate year for testing the endowment theory. The model relies on the specification of a long-run equilibrium without market distortions. It can hardly be argued that the economies of Europe and Japan were in an equilibrium in 1947; rather, they were beginning a process of rapid dynamic adjustment in production and factor supplies. Robert Baldwin recomputed Leontief's ratios for the United States using the 1958 input-output table and 1962 international trade data, with the result that the paradox was still strongly in evidence. However, the analysis by Robert Stern and Keith Maskus showed that by 1972, American exports had become capital-intensive relative to imports. Thus, the paradox may have been reversed by the 1970s. Even so, there remains the puzzling fact noted by Harry Bowen that in the post-War period, American relative capital abundance seems to have declined.[11] It is unclear how the relative capital intensity of exports could have risen in that context.

Alternative Explanations for the Paradox

A more revealing line of objections to Leontief's finding has come from noting that the assumptions of the Heckscher-Ohlin model are too strict to be believed. Indeed, the most enduring and valuable outcome of the debate over the Leontief paradox is that it stimulated trade theorists to think more fully about the implications of departures from those assumptions.

For example, it is possible that the international structure of trade barriers could partially explain Leontief's result. Our analysis of the endowment model showed that free trade would lower the real incomes of each country's scarce factor, providing an incentive for that factor to lobby for import protection. Thus, the United States might be expected to have high trade barriers to labor-intensive imports, and some foreign countries might erect restrictions on capital-intensive imports. These policies could reduce

the levels of trade below those expected from endowment-based comparative advantage. (Recall that trade restrictions cannot overturn comparative advantage altogether unless there are significant subsidies paid to exporting and importing the "wrong" commodities.) While this possibility is surely important, it has proved most difficult to test conclusively because tariffs have complicated effects across sectors in general equilibrium.[12]

Another possibility is that preferences across countries differ rather than being identical and homogeneous. Certainly, if there are significant *taste biases* in the sense that some countries have strong preferences for goods in which they would otherwise have a comparative advantage, the pattern of trade could be reversed. We noted in the last chapter that consumption patterns do vary markedly across nations. As we discuss later, this fact seems to help explain a significant share of world trade.

Another important possibility is that countries do not share access to identical technologies. This observation is at the core of the product-cycle model and related ideas about trade in a more dynamic context. The issue here is that trade data for a particular year may not reflect a long-run, static equilibrium so much as a short-run, dynamic transition under varying technologies. In the product cycle, for example, American exports of new goods may seem to be labor-intensive when, in reality, they make relatively heavy use of new technological information through the employment of highly skilled, technical labor inputs such as engineers and scientists.

The most substantive objection to Leontief's procedure is simply that it is inadequate to suppose that there are only two primary factors of production, capital, and labor, in the world. Various forms of land and natural resources also serve as sources of comparative advantage. Indeed, these may be the most relevant factors for the Heckscher-Ohlin model because they are internationally immobile. Further, physical capital and labor exist in different forms. We should not expect the average worker in higher-income countries to share identical productivity characteristics with the average worker in lower-income countries. Rather, national labor forces consist of different endowments of laborers of various skills, with skills being higher in countries that invest more in education and training. Because acquiring new skills often involves a lengthy investment process, it is reasonable to suppose that laborers of widely varying skills do constitute different endowments. For this reason, laborers are often distinguished by their *human capital,* or accumulated investments in education and training.

Regarding natural resources, Jaroslav Vanek made the observation that the Leontief paradox is consistent with the possibility that the United States was abundant in capital and labor but scarce in natural resources, such as oil, the extraction of which is quite capital-intensive, as we indicated in Chapter 8.[13] Thus, in the act of importing scarce natural-resource services, the United States was implicitly importing the services of capital as well. Some empirical economists accounted for this possibility in their work by excluding certain natural-resource-intensive products from their

computations. In the study by Stern and Maskus, for example, this exclusion reinforced their finding that the Leontief paradox was reversed in 1972. In general, relative supplies of land and natural resources clearly affect world trade patterns. The United States and Canada are both large net exporters of agricultural commodities, while Canada also exports the services of energy and minerals. In contrast, Japan, Hong Kong, and Singapore are extremely scarce in such resources, and their imports are dominated by oil, raw materials, and food.

With respect to labor skills, Donald Keesing was the first to show that a disaggregation of U.S. labor by skill type was important in explaining the factor contents of trade.[14] American net exports were clearly intensive in highly skilled laborers, such as professional workers. This finding has proved to hold strongly in examinations of other data sets and definitions of human capital. Most trade economists agree that a fundamental determinant of U.S. comparative advantage is a relatively abundant supply of highly-skilled labor. This finding explains the fact that average wages in American export sectors are markedly higher than average wages in import-competing sectors. This is true of other developed nations as well.

These observations were important in spurring additional theoretical research, especially in terms of the product cycle and related ideas. However, the role of additional factors is properly incorporated into the factor-endowments model only through the factor-content (or HOV) theorem, which we discussed in Chapter 8. We simply state the mathematical result here, because it is useful in motivating subsequent discussion, and we refer interested readers to an article by Edward Leamer for a formal proof.[15] Retain all assumptions of the Heckscher-Ohlin theorem, except allow there to be m factors and n goods, with $n \geq m$. For any country j, let s^j be its proportion of world income. Finally, let F_i^j and F_i^w denote the endowments of factor i for country j and the world, respectively. Let us define a ranking of factor abundance and scarcity for country j by virtue of its share of world endowments of each factor:

$$(F_1^j/F_1^w) > (F_2^j/F_2^w) > \cdots > (F_i^j/F_i^w) > s^j > (F_{i+1}^j/F_{i+1}^w) > \cdots > (F_m^j/F_m^w) \qquad (14.1)$$

Under the HOV theorem, country j's consumption share lies somewhere in the middle of this chain, as we have written. It also follows that country j's net exports of the services of any factor are positive if its abundance ranking for that factor lies above the consumption share, and its net exports are negative if its ranking lies below that share. That is, a country exports the services of its abundant factors and imports the services of its scarce factors when factor abundance is measured relative to a global standard. This is the essence of the factor-content theorem.

The practical importance of this theorem is that it can be used to rank a country's factor endowments in terms of their relative abundance as revealed by trade. Indeed, Leamer used Leontief's own data to show that the United States had in fact been revealed to be relatively abundant in capital and scarce in labor with the allowance of more than two factors. Thus, there may never have been a Leontief paradox. Moreover, once these rankings are compiled, they may be compared to actual data on national and world factor endowments as a full test of the HOV theorem. Such tests have been performed for the United States by Keith Maskus and for a collection of countries by Harry Bowen, Edward Leamer, and Leo Sveikauskas.[16] The results of these tests indicate that the rankings of factor abundance revealed by trade data do not convincingly replicate actual endowment rankings, suggesting that the theorem does not hold literally in real-world situations.

Judging the Importance of Factor Endowments in Explaining Trade

Of course, we should not be surprised by the inability of actual data to conform to the rigorous assumptions of the HOV model. These assumptions are made simply to study the theoretical effects of endowment differences on trade in factor services. No one would be surprised to learn that the world is not perfectly competitive and that there exist market distortions, such as taxes and trade barriers. Rather than testing the truth of a theory based on unrealistic assumptions, it is more illuminating to measure how strongly factor endowments seem to affect actual trade flows.

Two statistical approaches to this question have been employed. The first technique is to relate measures of U.S. net exports across manufacturing industries to factor intensities in production. Two studies worth mentioning are one by Robert Baldwin and one by Robert Stern and Keith Maskus.[17] Baldwin found that capital intensity was negatively related to net exports, another version of the Leontief paradox, but that human capital was positively associated with net exports. Stern and Maskus demonstrated that over the last few decades, U.S. comparative disadvantage became more strongly associated with lower-skilled labor inputs. They further showed the importance of human capital and technological inputs, such as research and development spending, in explaining the structure of U.S. trade. In general, the results from studies of this nature are consistent with economists' beliefs about underlying relative endowments across countries, providing indirect support for the Heckscher-Ohlin theory.

A second approach, developed by Edward Leamer in an important book, is quite close to the spirit of the Heckscher-Ohlin-Vanek model.[18] In that model, if we assume that the numbers of factors and goods are equal, that factors are fully employed and homogeneous in all countries, and that factor prices are equalized internationally, the net exports of any commodity are a (complicated) linear function of the difference between a country's vector of factor supplies and its vector of factor demands. Assuming identical and homothetic tastes, the demand for factors is simply the product of the country's global income share and global factor endowments.

TABLE 14.1
Relative factor endowments, 1975 (percent of world endowments)

Factor*	Canada	Fed. Rep. Germany	India	Japan	Mexico	South Korea	U.K.	U.S.
GNP	3.46	9.22	1.99	11.14	1.00	0.47	5.10	32.94
Capital	3.65	10.28	1.18	14.92	0.70	0.37	4.87	29.36
Prof. Labor	2.37	6.11	13.47	8.03	1.64	0.70	6.00	24.53
Lit. Labor	1.74	5.19	14.29	11.05	2.35	1.86	4.77	17.11
Illit. Labor	0.02	0.07	65.24	0.20	1.30	1.41	0.06	0.15
Trop. Land	0.00	0.00	9.11	0.00	4.02	0.00	0.00	0.14
Dry Land	2.38	0.00	5.68	0.00	6.24	0.00	0.00	32.44
Temp. Land	38.03	1.27	3.77	1.91	0.00	0.51	1.25	18.47
Coal	1.96	10.80	8.49	1.67	0.32	1.35	11.13	50.87
Minerals	19.06	1.79	3.43	1.89	3.09	0.31	0.98	27.58
Oil	9.76	1.25	0.73	0.16	2.18	0.00	1.63	58.04

Source: E. E. Leamer, *Sources of International Comparative Advantage: Theory and Evidence.*

*Prof. Labor: professional and technical workers; Lit. Labor: literate, nonprofessional workers; Illit. Labor: illiterate workers; Trop. Land: land in tropical rainy climate; Dry Land: land in dry climate; Temp. Land: land in humid, temperate climate.

It is possible to use a statistical procedure termed *regression analysis* to estimate this linear relationship. Regression analysis involves finding the linear equation that most closely fits a set of data that has observations on a dependent variable (here, net exports in each commodity) and one or more explanatory variables (here, national endowments of factors). For each of ten broadly defined commodities, such as raw materials and labor-intensive manufactures, Leamer estimated this relationship using a sample of 47 countries (defined to make up the "world,' this sample did cover most of the major trading nations) with measures of 10 factor endowments in 1975. The regression estimates were then taken as measures of the link between endowments and trade.

It is worth mentioning some basic results. Table 14.1 shows computations of relative factor endowments, or shares of each country in the world supply of factors, for selected nations. The entries in the first row are national shares in world GNP, or the s^j-terms in our earlier analysis. The remaining rows show each country's share in the world supply of factor endowments. In theory, these may be compared to the GNP shares to measure relative factor abundance. For example, in 1975 Japan had 14.92 percent of the world capital stock, which exceeded its GNP share of 11.14 percent, suggesting that Japan was capital-abundant. Japan was clearly very scarce in land and natural resources. The United States was abundant in dry land, coal, and oil and scarce in lower-skilled labor and tropical land. In fact, the United States was a large net importer of oil, suggesting that its tastes are biased heavily toward oil consumption in comparison with other countries.[19] Canada was abundant in temperate land, minerals, and capital.

India seemed to be scarce in capital and oil but relatively abundant in other factors. The fact that India was well-endowed with so many factors but had a small share of world GNP reflects the very low productivity of inputs there. This suggests that production functions in India were significantly inferior to those in other countries in 1975, in contrast to the Heckscher-Ohlin assumption of identical technologies.

The interesting question is how these endowments relate to trade flows. Leamer's preferred regression estimates of the trade equations are shown in Table 14.2 for selected groupings of commodities. Each entry in the table indicates the change in net exports, in thousands of U.S. dollars, that is estimated to result from a unit increase in the associated endowment. Thus, a $1 million increase in the capital stock of the average country would reduce net exports (or raise net imports) of raw materials by $8,800 and of cereals by $4,300. However, it would increase net exports (or lower net imports) of animal products by $40, of labor-intensive products by $1,000, of capital-intensive products by $16,500, of chemicals by $3,800, and of machinery by $29,100. Readers may verify that the results make good sense. Increases in land and resource supplies tend to expand exports of raw materials and agricultural goods while reducing exports of manufactured goods. Expansions of the stock of literate labor (essentially educated, blue-collar workers) raise net exports of most manufactured goods but reduce net exports of materials and agricultural commodities. Professional labor is a source of comparative advantage for chemicals. Overall, the equations fit the data well; each equation typically explains as much as 50 to 60 percent of the observed trade patterns. The conclusion from this analysis is that

TABLE 14.2
Estimates of the effects of factor endowments on trade, 1975

	Commodity group*						
Factor	Raw Mtls	Animal	Cereals	Labor-Int.	Capital-Int.	Chemicals	Machinery
Capital	−8.8	0.04	−4.3	1.0	16.5	3.8	29.1
Prof. labor	303.1	−279.4	946.3	−699.7	−1947.9	481.7	−1177.4
Lit. labor	−59.4	−17.3	−97.4	78.9	126.7	−53.4	77.7
Illit. labor	2.5	17.9	−18.8	4.8	39.1	−4.4	8.3
Trop. land	−0.1	−0.3	2.3	−0.5	−0.8	−0.8	−0.7
Dry land	−0.3	0.7	1.0	−0.3	−0.3	−0.2	−0.3
Temp. land	0.6	7.4	20.6	−3.8	−11.8	−8.5	−19.2
Coal	0.4	−0.05	0.03	−0.07	−0.08	0.03	−0.05
Minerals	1.1	−0.03	−0.01	−0.05	−0.07	−0.05	−0.08
Oil	0.04	0.03	0.24	−0.04	−0.2	−0.04	−0.2

Source: E. E. Leamer, *Sources of International Comparative Advantage: Theory and Evidence.*

*Raw mtls.: raw materials; Animal: animal products; Labor-int.: labor-intensive manufactures; Capital-int.: capital-intensive manufactures. The unit of measurement for each commodity group is $1000. The unit for each type of labor is 1000 workers; for each type of land 1000 hectares; for coal, minerals, and oil $1000; and for capital $1 million.

factor endowments are an extremely important influence on comparative advantage in world trade.[20]

14.5 PREFERENCES, TECHNOLOGY, AND SCALE ECONOMIES: THE IMPORTANCE OF INTRA-INDUSTRY TRADE

As we have emphasized in this book, there are other theoretical determinants of trade than Ricardian technological differences and variations in factor endowments. Prominent in this discussion have been differences in tastes as in the Linder Hypothesis, dynamic evolution of comparative advantage as in the product-cycle model, and increasing returns to scale and imperfect competition. These views have been subject to considerably less empirical testing than the endowment model. Again, rigorous tests of these ideas are difficult to execute. Nonetheless, we can gain some insights from a brief review of the major studies available. We proceed by considering work that is focused on each issue separately. However, one feature these theoretical notions have in common is that each can serve as an explanation for the existence of intra-industry trade (IIT). We discuss the empirical significance of IIT in the final subsection.

International Preferences

Casual consideration of international trade data would suggest that the Linder hypothesis provides a powerful explanation of trade patterns. As we show later, the bulk of international trade in manufacturing goods takes place among the developed countries, which have similar per capita incomes. Thus, it seems that a regression in which bilateral trade flows are explained by differences in per capita incomes (with the amount of trade rising as incomes grow more similar) should be successful. Indeed, early exercises of this sort consistently found such a correlation. However, countries with similar per capita incomes also tend to be geographically concentrated, such as those in Western Europe. Thus, trade among these countries may be accounted for simply by lower transportation costs. For example, this view is consistent with one explanation of the failure of Japan to engage in as much IIT as other developed countries: Japan is far-removed from other industrial markets. One test of this hypothesis is to relate *changes* in bilateral trade and per capita incomes over time, which holds geographical proximity (and relative transport costs) fixed. Several efforts along these lines have found no clear evidence for the Linder hypothesis.[21]

One problem with the Linder hypothesis is that it places relatively little theoretical structure on the determinants of trade and therefore provides little guidance to the empirical analyst. A more recent and very promising line of research makes assumptions about the form in which tastes are allowed to differ. As we discussed in Chapter 13, if tastes are

quasi-homogeneous, a model may be developed to help explain the observed international variations in the shares of consumption of different goods. We noted earlier that the proportion of income spent on food declines rapidly with the level of per capita income, falling from an average of 50 percent for the poorest countries to 17 percent for the richest countries. Thus, food is clearly a *necessity good* in economic terms. In contrast, relative expenditures on medical care, transport and communication, and education tend to rise sharply with incomes. Linda Hunter has explored the implications of quasi-homogeneity for trade patterns.[22] She estimated demand functions for 11 commodity groups across a sample of 34 countries in 1975 based on a particular utility function that implies quasi-homogeneous demand functions. She then "neutralized" the effects of demand differences by forecasting what trade would be if preferences were homogeneous and all markets continued to clear. The difference in these estimates suggests the importance of nonhomogeneity in explaining world trade. Her results suggest that as much of 29 percent of world trade may be caused by nonhomogeneous preferences, with this effect being stronger for higher-income countries. Hunter's approach has been criticized because of the special nature of the underlying utility functions, and her results should not be considered definitive. Nevertheless, the evidence is most intriguing and suggests that empirical analysts should devote more effort to the issue of tastes and trade.

The Product Cycle

The product cycle is an intuitively appealing notion that is supported by substantial anecdotal evidence. For example, virtually any product in the consumer electronics industry, such as color televisions, video recorders, and semiconductors, seems to obey the prescriptions of the model. New innovations in the technologies underlying these products are predominantly developed in the industrial nations while, with little delay, mass production is located in the developing economies. Similar processes seem to characterize chemicals, textiles, and some other goods. Indeed, a number of industry case studies, now rather out of date, have yielded results that support the model.

Despite this fact, generalized testing of the product cycle has been frustrated by contradictory results. Again, a primary difficulty is distinguishing this theory from other possible explanations of trade. One clear example is that in numerous studies of American trade structure, there is a strongly positive correlation between net export strength by industry and the industry's investment in research and development (R&D). This seems to constitute strong indirect evidence for the product-cycle model, except that it is quite possible that R&D spending is simply a surrogate measure for the relative abundance in the United States of technically skilled labor inputs. The latter factor may well be the ultimate source of American comparative

advantage. Further, there always remains an issue of underlying causality—is high R&D spending a determinant of trade or the result of profitable trading opportunities based on other sources of comparative advantage? This question has yet to be answered definitively.

One direct approach to measuring the product-cycle model, pioneered by Gary Hufbauer, is the practice of identifying the first date at which a particular product entered world trade. Countries like the United States might be expected to have strong net export positions in newer goods. This proposition has received only weak support in the literature. More recently, David Audretsch argued that it should be possible to characterize the newness of goods in terms of their underlying production profiles, with global production of a good rising at an increasing rate for innovative goods and remaining fairly stable for standardized goods.[23] His correlations concurred somewhat with this hypothesis, suggesting that innovativeness may be a measurable concept in world trade.

The product-cycle model inherently concerns the dynamics of trade. It helps explain theoretically the existence of multinational enterprises (MNEs), among other things. As we discussed in the last chapter, the practice by MNEs of trading inputs and outputs within the firm and rapidly relocating production has probably reduced the length of the product cycle to an indistinguishable length. In this sense, a proper test of the model would take into account the existence and growth of MNEs, attempting to explain the reasons for their existence across industries. To date, this task has not been adequately performed. Further, we noted the importance to the model of understanding how rapidly new technological information diffuses among countries through technology transfer, product copying, and the like. This process also has not yet been studied in a satisfactory way.

We must conclude that adequate formal testing of the product-cycle model has yet to be designed and performed. The model certainly helps explain some features of the world economy that most economists believe are important, such as the need to invest in product development and quality and the desire to shift production location over time. Further, it may be that the very existence of MNEs is indirect validation of the model.[24] At this time, however, we remain fairly skeptical of the model's applicability to the major proportion of world trade.

Scale Economies and Imperfect Competition

Most economists would regard economies of scale as a significant factor influencing international trade. An obvious point to make in this context is that international trade is dominated by transactions among very large firms, rather than by the perfectly competitive, atomistic firms of standard trade theory. Indeed, recent research has demonstrated that a significant proportion of global trade in manufactures takes place as internal transactions within multinational enterprises.[25] These facts would be of little theoretical

interest if such firms existed simply to conduct trade based on traditional notions of comparative advantage. However, our earlier discussion pointed out that the existence of increasing returns can have important and distinctive effects on trade and the gains from trade. This issue has commanded surprisingly little empirical work despite its obvious importance.

That increasing returns exist at the level of individual plants in manufacturing is clear.[26] However, direct evidence that scale economies affect trade tends to be contradictory.[27] Many studies in numerous countries have detected a positive correlation between the size of industrial plants and production for exports, strongly suggesting that scale economies facilitate the achievement of lower costs for exporting. However, when analysts add a variable measuring economies of scale to statistical studies of trade flows based on theoretical models, the results provide only weak evidence that economies of scale contribute to international trade. It should be noted that measuring scale economies in a consistent way across industries has proven an elusive task. For example, standard measures of how concentrated an industry is, such as the percentage of industry sales accounted for by the largest four firms, are not necessarily related to significant scale economies or to barriers to entry. More fundamentally, we have pointed out that increasing returns may occur in a variety of important ways. Besides plant size are issues of geographical concentration, cost-reducing innovations in intermediate inputs, diffusion of technological information, and learning economies from repeated production. Indeed, evidence is emerging that the use of new intermediate goods, such as computers and advanced machinery, significantly improves the productivity of manufacturing firms and helps expand exports. However, most of these impacts—arguably the most important influences on scale economies and trade—have not been measured particularly well.

Such problems have pushed economists toward two alternative and more revealing approaches. One body of research is termed *numerical general equilibrium analysis*. In this approach, economists develop models of an entire economy (or, say, several economies linked together in a regional trade agreement), taking account of various market structures in different industries. A key feature of these models is that they are capable of incorporating important economy-wide constraints, such as a fixed supply of labor and capital in the short run, that would have significant influences on the degree to which sectoral outputs can change in response to a shift in government policy. Another feature is that they can provide estimates, or even well-informed guesses, about the values of important parameters in each industry, such as the ease of technical substitution among inputs for production functions, elasticities of demand for outputs and inputs, and the extent of differentiation of products coming from different countries. Typically these estimates are taken from other published studies and combined with a set of data on outputs, consumption, and trade for a particular year, assuming the available data represent a general equilibrium for the

economy. If the model can be combined with the data in a way that does not violate any economic constraints and that is consistent with fundamental economic behavior (that is, consumers maximize utility, producers maximize profits or minimize costs, and the government attempts to provide public services in an efficient way), it can be used for policy analysis. Specifically, the effects of changes in trade policy (such as tariff reductions or entrance into a free trade area) on industry outputs, trade, and consumer welfare may be simulated under the given set of parameters. It is critical that various ranges of key parameters be used to see how sensitive the results are to particular assumptions about their values.

For our purposes, a key set of parameters is the importance of scale economies in various sectors. Models in which economies of scale are presumed to exist in manufacturing industries tend to show significant interrelationships between trade and returns to scale. For example, Richard Harris constructed a numerical general equilibrium model in which the manufacturing sector was characterized by increasing returns to scale in order to examine the issue of how trade liberalization would affect the Canadian economy.[28] In this model, the bilateral removal of tariff and nontariff barriers in the context of a free trade area with the United States should allow Canadian industries to take advantage of scale economies as they sell in the larger U.S. market. The Canadian economy should also become more efficient because of industry rationalization, gains in productivity, shifts in resources among sectors, and price competition, as we have discussed in earlier chapters on imperfect competition. Harris' model predicts that, in the long run, the effects of the Canada-United States Free Trade Area should include a rise of 9 percent in Canadian real wages, 3.2 percent in Canadian real GNP, 8 percent in Canadian labor productivity, and 30 percent in real trade volumes. While these estimates are subject to some uncertainty, they clearly point out the large potential effects of trade liberalization in the presence of increasing returns. In contrast, regression estimates in static models with constant returns rarely discover a potential gain in real GNP from tariff cuts of more than 1 percent. This work has helped stimulate a large body of numerical general equilibrium models of various economies. These models have consistently pointed to the key role that scale economies play in shaping trade patterns and in achieving efficiency gains from trade.

A second approach has been to incorporate imperfect competition, presumably based on increasing returns, directly into trade models and to estimate or simulate its effects. An excellent review of this literature was provided by David Richardson.[29] Many of these studies have been concerned with oligopolistic competition in a particular industry and therefore, they say little about the general determinants of trade. Nonetheless, imperfect competition clearly plays a significant and perhaps dominant role in explaining trade in automobiles, aircraft, semiconductors, and computers, among other sectors.

Intra-Industry Trade

The three types of models just examined share an important characteristic. Each is capable, in theory, of explaining the prevalence of intra-industry trade (IIT) in the world economy. As we discussed in Chapter 13, this phenomenon is the simultaneous import and export by a country of very similar types of products. For example, the Linder hypothesis and related notions about trade among nations with similar per capita incomes is consistent with the observation that many slightly differentiated consumer goods are traded internationally. If trade takes place to satisfy tastes for variety in consumption, we can understand the fact that the United States and Canada have extensive two-way trade in beers, that French, Italian, and German wines are traded simultaneously with American wines, and that high-fashion clothing of different brands and styles is exchanged among high-income economies. Countries might also simultaneously import and export similar products that embody different levels of technological sophistication, as suggested by the product cycle models. Finally, scale economies allow firms in different nations to produce at lower cost by specializing in similar types of goods or inputs, with international trade among the countries serving as the incentive for that specialization. This process is characteristic of automobiles, machinery, chemicals, and other industrial commodities.

Intra-industry trade is important enough to warrant consideration of some details on its measurement. Some central facts about world trade in goods are presented in Table 14.3. In 1990, total world merchandise exports amounted to $3.188 trillion. It is interesting that the high-income, developed economies accounted for 74.6 percent of this trade. This is hardly surprising because these economies also produced 71.7 percent of world GDP. More striking is the fact that the developed countries traded extensively among themselves, accounting for $1.816 trillion or 57.0 percent of world exports.

TABLE 14.3
The composition of world merchandise exports by country group, 1990

Country group	Exports ($billions)	Share of world exports (%)	Export share in GDP (%)	Exports per capita
World	3188	100.0	14.3	$603
High-income developed	2379	74.6	14.9	$3062
Intra-HID	1816	57.0	11.4	$2337
Other high income	177	5.5	54.7	$4414
Upper middle income	307	9.6	20.2	$670
Lower middle income	184	5.8	19.8	$293
Low income	141	4.4	15.4	$46

Sources: World Bank, *World Development Report, 1992* and International Monetary Fund, *Direction of Trade Statistics Yearbook, 1993.*

Exports are also scaled in the table by GDP and population in order to help neutralize the effects of country size on these comparisons. Thus, the high-income, developed economies tend to have somewhat lower export-to-GDP ratios (though there are marked exceptions; for example the Swiss ratio is 28.4 percent). At the same time, their populations are highly productive in exports, with over $2000 per person exported among them. Poor countries engage in far fewer exports per capita and claim a far smaller share of world trade.

The dominance of world trade by exchange among the developed economies is difficult to reconcile with standard Heckscher-Ohlin and Ricardian trade models. The Heckscher-Ohlin model, for example, predicts that trade should transpire largely between countries with markedly different relative factor endowments. Thus, we would expect *inter-industry* trade in very different products to dominate world trade. To be sure, such trade is important in several commodities and for several countries. Japan must import a disproportionate share of the world's natural resources because of its scarcity of land and minerals. Developing countries in Asia export labor-intensive manufactures, such as clothing, in return for machinery and specialized inputs. We suggested earlier in this chapter that the factor-content theorem provides a credible explanation for an important component of world trade.

Nonetheless, the significant degree of trade among the developed economies, which have relatively similar factor endowments and technologies, must be caused primarily by other economic factors. Our answer lies in extensive amounts of IIT that take advantage of economies of scale and differentiation of products. Table 14.4 provides some simple statistics on IIT for certain countries and products. The formula for computing these IIT indexes is the following:

$$IIT = 100[1 - (|e_j - i_j|)/(e_j + i_j)] \quad (14.2)$$

TABLE 14.4
Intra-industry trade by commodity and country, 1990

Commodity	Canada	Fed. Rep. Germany	Japan	South Korea	U.K.	U.S.
Fuels	73.1	31.8	4.7	11.9	99.9	30.7
Chemicals	92.8	75.8	99.0	50.6	89.9	75.0
Special industrial machinery	62.5	45.6	35.5	26.9	89.7	91.8
Computers	46.3	74.7	39.0	68.4	95.3	99.8
Automobiles	79.7	58.7	26.0	10.1	62.3	37.5
Clothing	24.1	52.1	0.2	0.2	60.7	17.3
Precision instruments	48.9	71.6	70.4	37.6	91.6	67.7

Source: Computed from United Nations, *Yearbook of International Trade Statistics, 1990.*

This index, originally developed by Herbert Grubel and Peter Lloyd, can range from zero, indicating complete inter-industry trade in industry j (that is, either exports or imports are zero) to 100, indicating complete intra-industry trade (exports equal imports).[30] One significant problem exists with these computations. When IIT indexes are calculated for aggregated commodity groups, as in Table 14.4, they tend to overstate the true amount of IIT by lumping together goods that really are not very similar. It is more appropriate to compute them for highly detailed and disaggregated commodities, which would have the effect of considerably reducing the figures in Table 14.4. Nonetheless, studies that have done this consistently find that IIT is an important phenomenon across a wide range of goods and countries.

Note in Table 14.4 that Germany and the United Kingdom tend to have especially high IIT measures. This reflects their membership in the European Economic Community, where proximity and common trade practices act to encourage considerable trading of similar commodities across borders. In particular, notice the high percentages of IIT in clothing trade for these countries. There is a great deal of cross-border trade in fashion in Western Europe. Similarly, Canada has high indexes reflecting its overwhelming trade relationship with the United States. Japan, however, tends to have smaller IIT figures, which, as we suggested earlier, reflects both its peculiar land scarcity and its long distances from other developed countries. In this regard, Japan is more like Korea than it is like the United States.

At the same time, however, IIT varies across commodities. Chemicals, computers, industrial machinery, and precision instruments tend to have a high percentage of IIT in the industrial economies. All of these products are subject to technological differentiation and scale economies. Canada has a very high IIT figure in automobiles, which is caused by substantial two-way trade with the United States. These two countries share largely free trade in automobiles as a result of The United States-Canada Automotive Agreement of 1965. Again, IIT in automobiles is prevalent in Western Europe. For most countries IIT is relatively small in clothing, which is an indication that trade in that sector is largely driven by comparative advantage (Korea) and disadvantage (Canada, Japan, and the United States) in labor-intensive goods. Fuels are also largely governed by inter-industry forces.

These highly aggregated figures are crude, yet they suggest important phenomena in world trade. Numerous economists have attempted to use statistical techniques to explain the structure of IIT across commodities and countries. One classic reference is the study by Bela Balassa.[31] Balassa demonstrated that trade in manufactures increased rapidly in Western Europe after the formation of the European Common Market, indicating that forces such as product differentiation, scale economies, and industrial rationalization tend to accompany economic integration activities among similar countries. Similarly, in their book, Grubel and Lloyd were the first

to carefully document the existence of IIT. They used their findings to make a series of hypotheses about the sources of IIT. It is interesting to note that IIT is one economic phenomenon that was measured empirically before it was explained theoretically. In this sense, the Grubel and Lloyd study is one of the most important empirical efforts in international trade because it stimulated substantial theoretical inquiries into the relationships among imperfect competition, demand, and trade.

In a recent study, Balassa attempted to capture basic determinants of IIT indirectly by measuring certain *national* characteristics rather than *commodity* and *industrial* characteristics, such as product differentiation and scale economies.[32] He hypothesized that IIT would rise with the level of per capita GNP (a measure of economic development and also an indirect indicator of product differentiation) and the level of GNP itself (a measure of market size and scale economies). He also expected IIT to fall as geographic distance between trading partners became larger and as trade restrictions became more significant. Finally, he made allowance for the fact that some countries share borders and some countries are members of economically integrated areas, such as the European Community. These latter variables would be positively correlated with IIT to the extent that they allowed for greater information flows about differentiated products and allowed for rationalized industrial structures. Balassa analyzed 38 developed and developing countries in 1971, using disaggregated trade data to compute an IIT index for each country. His results seemed to support his hypotheses, as each variable performed in the expected direction and was statistically significant. Particularly important factors seemed to be membership in the European Community, the existence of common borders, and the level of economic development.

Studies such as Balassa's clearly suggest that IIT is an important phenomenon and that it is caused by identifiable factors. Nevertheless, these studies attract criticism of the kind we discussed earlier. For example, no one has yet specified a fully integrated theory of IIT that would allow us to determine appropriate equations for use in these studies. In that sense, the results are difficult to interpret, and indeed, they seem to change considerably depending on the equations used and the variables included. Also, the use of broad aggregate variables as loose and indirect measures of important characteristics underlying IIT is questionable. It cannot be stated with confidence that per capita GNP measures product differentation and tastes as opposed to, say, relative capital and labor endowments. In fact, per capita GNP has often been used as a variable measuring factor endowments in studies of the Heckscher-Ohlin model. Accordingly, the status of empirical work on the endowment model is generally equivalent to the status of empirical work on models of IIT. Both are considered important determinants of actual trade flows and should be viewed as complementary explanations of international commerce.

14.6 CONCLUDING REMARKS

Active research continues to be done on empirical testing of international trade theory. It has proven difficult to observe practical versions of our complicated general equilibrium trade theories. Because of these difficulties, economists have had more success in measuring the contributions of various determinants of trade flows. Economists generally agree that relative labor productivity, factor endowments, tastes, and product differentiation with scale economies all seem to provide important sources of comparative advantage. While it is not possible to determine the exact contribution of each influence, our conclusion is that there are numerous important variables that clearly induce countries to trade among themselves.

The main points for review from this chapter include the following.

1. There is substantial evidence that economies garner significant gains from trade. One study of Japan in the 19th century actually demonstrated such gains in a comparison of autarky with liberalized trade. Indirect evidence of this point is also strong, as it is clear that relatively open economies have grown more rapidly than relatively closed economies in the post-war period.
2. Relative labor productivities are correlated with relative export performances of nations, as suggested by the Ricardian model. However, these results are difficult to interpret in light of the fact that other theories predict similar outcomes.
3. The Leontief paradox was the unexpected finding that American exports were labor-intensive and imports were capital-intensive in 1947. This result was tremendously important in stimulating further thinking about the determinants of trade. Attempts to explain the finding through relaxing certain assumptions of the Heckscher-Ohlin model have met with mixed success. More importantly, economists now believe that numerous factor endowments, including capital, types of labor, and natural resources provide sources of comparative advantage. This perspective has led to the development of the factor-content trade theorem, which is true under very general circumstances.
4. The rigorous assumptions of the factor-endowments model have forced economists to attempt to measure the contribution of endowments to trade rather than test the theorem literally. This effort has proved largely successful, pointing to the importance of factor endowment differences in actual trade flows. Generally, differences in natural resources and land supplies provide important sources of comparative advantage, as do differences in labor skills.
5. Differences in tastes seem to have a significant influence on world trade flows, though research into this question is still relatively new. Broad empirical evidence in favor of the product-cycle model has not yet been found.

6. Models of IIT point toward other influences on trade, including product differentiation, imperfect competition, and scale economies. While IIT is certainly a measurable and important phenomenon, no full "test" of its theoretical predictions yet exists. However, economists have tried, with substantial success, to measure the contribution of its underlying determinants to the extent of IIT. Factor endowments as well as characteristics of IIT provide important explanations for the structure of international trade, with the former being more closely associated with trade in primary goods and labor-intensive manufactures and the latter being more closely associated with trade in differentiated manufacturing goods.

PROBLEMS

1. Explain why Leontief's result was considered to be paradoxical. Why do economists consider his approach to have been only an incomplete "test" of the Heckscher-Ohlin model?
2. Criticize the various theoretical attempts to reconcile or explain the Leontief paradox. Further, explain how other theories of trade could help explain the paradox.
3. Why do countries engage in intra-industry trade? List several consumer commodities for which both domestic and imported varieties are available for purchase. Do you think that trade in such goods can be reconciled with the notion that trade is determined by differences in factor endowments? Differences in technologies? Why or why not?
4. The following data represents hypothetical trade in four commodities for some country. Compute the IIT index for each commodity. Which commodities are more characterized by inter-industry trade and which by intra-industry trade? How would you suggest using these data to calculate an IIT index for the country as a whole?

	Exports	Imports
Commodity 1	$0	$5000
Commodity 2	$3000	$2500
Commodity 3	$10000	$1000
Commodity 4	$4500	$7000

5. Below are actual trade data (in millions of dollars) for the United States in 1989 for a broad category of machinery and some detailed subcategories. Compute the IIT indexes at each level. Why do you suppose the indexes for the more detailed categories are lower? What would you conclude about the accuracy of computing IIT indexes directly from aggregate categories as opposed to taking an average of the indexes from the subcategories that comprise them?

	Exports	Imports
General industrial machinery	$13095	$14974
Certain fuel pumps	$115	$345
Ball bearings	$106	$574
Pressure-reducing valves	$48	$26
Forklift trucks	$315	$737

NOTES

1. Interested readers are referred to the excellent survey by Deardorff (1984).
2. Note that this is a common problem in the social sciences, in which it is largely impossible to isolate the effects of a particular variable on some system as may be done in a chemistry laboratory. Recently, economists have begun constructing laboratory experiments with human and animal subjects to investigate relatively simple questions in economic theory. However, this approach is virtually absent from the testing of international trade theories.
3. See J. Richard Huber (1971).
4. We may currently be witnessing a similar transformation as the economies of Eastern and Central Europe and China reorient themselves toward greater trade with the Western market economies.
5. See the paper by Sebastian Edwards (1993) for a review.
6. The exceptions are countries for which economic activity is dominated by the production of a primary commodity, such as oil or cocoa, for export. Notice, however, that comparative advantage in such goods is surely related more to endowments of natural resources than to labor-productivity advantages.
7. See the study by G.D.A. MacDougall (1951).
8. See Robert M. Stern (1962).
9. See W. Leontief (1953).
10. See P. Heller (1976).
11. See R. E. Baldwin (1971), R. M. Stern and K. E. Maskus (1981), and H. P. Bowen (1983).
12. Interested readers are referred to the article by Edward Leamer (1988).
13. See J. Vanek (1963).
14. See the article by D. B. Keesing (1965).
15. See E. E. Leamer (1980).
16. See K. E. Maskus (1985) and H. P. Bowen, E. E. Leamer, and L. Sveikauskas (1987).
17. See R. E. Baldwin (1971) and R. M. Stern and K. E. Maskus (1981).
18. See E. E. Leamer (1984).
19. In reality, the U.S. share of the world's oil is not this large because the data sample excludes many nations with large stocks of petroleum, such as those of the Middle East.
20. For more recent estimates of world endowments and trade equations, see the paper by K. E. Maskus (1991).
21. However, in his 1990 article, Jeffrey Bergstrand has provided firmer evidence for the role of preference similarity in explaining bilateral trade flows.
22. See the article by L. Hunter (1991).
23. See the papers by G. C. Hufbauer (1970) and D. B. Audretsch (1987).
24. There are numerous possible explanations for MNEs, as we discuss in Chapter 22.
25. For example, in a study of U.S. trade flows with 27 countries in 1989, S. Lael Brainard (1993) found that the share of both U.S. imports and exports accounted for by intra-firm transfers was around 25 percent.
26. See, for example, the review in F. M. Scherer (1980).
27. An excellent review is provided by James R. Tybout (1993).
28. See R. G. Harris (1991). The structure of the model is explained by D. Cox and R. Harris (1985).
29. See the article by J. D. Richardson (1989).
30. See H. G. Grubel and P. J. Lloyd (1975).
31. See B. Balassa (1966).
32. See B. Balassa (1986). There are now hundreds of such studies in the literature.

REFERENCES

Audretsch, D. B. (1987). "An Empirical Test of the Industry Life Cycle." *Weltwirtschaftliches Archiv* 123: 297–308.

Balassa, B. (1966). "Tariff Reductions and Trade in Manufactures among Industrial Countries." *American Economic Review* 56: 466–473.

———.(1986). "Intra-Industry Specialization: A Cross-Country Analysis." *European Economic Review* 30: 27–42.

Baldwin, R. E. (1971). "Determinants of the Commodity Structure of U.S. Trade." *American Economic Review* 61: 126–146.

Bergstrand, J. (1990). "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade." *Economic Journal* 100: 850–868.

Bowen H. P. (1983). "Changes in the International Distribution of Resources and Their Impact on U.S. Comparative Advantage," *Review of Economics and Statistics* 65: 402–414.

Bowen, H. P., Leamer, E. E., and Sveikauskas, L. (1987). "Multicountry, Multifactor Tests of the Factor Abundance Theory." *American Economic Review* 77: 791–809.

Brainard, S. L. (1993), "An Empirical Assessment of the Factor Proportions Explanation of Multinational Sales," National Bureau of Economic Research working paper.

Cox, D., and Harris, R. (1985). "Trade Liberalization and Industrial Organization: Some Estimates for Canada." *Journal of Political Economy* 93: 115–145.

Deardorff, A. V. (1984). "Testing Trade Theories and Predicting Trade Flows." In R. Jones and P. Kenen, eds., *Handbook of International Economics: Volume I.* Amsterdam: North-Holland, 467–518.

Edwards, S. (1993). "Openness, Trade Liberalization, and Growth in Developing Countries." *Journal of Economic Literature* 31: 1358–1393.

Grubel, H. G., and Lloyd, P. J. (1975). *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products.* New York: John Wiley.

Harris, R. G. (1991). "Symposium—The Canada-U.S. FTA: Economic Impact and Transition Effects." *Journal of Policy Modelling* 13: 421–434.

Heller, P. S. (1976). "Factor Endowment Changes and Comparative Advantage: The Case of Japan, 1959–1969." *Review of Economics and Statistics* 58: 283–292.

Huber, J. R. (1971). "Effect on Prices of Japan's Entry into World Commerce after 1858." *Journal of Political Economy* 79: 614–628.

Hufbauer, G. C. (1970). "The Impact of National Characteristics and Technology on the Commodity Composition of Trade in Manufactured Goods." In R. Vernon, ed., *The Technology Factor in International Trade.* New York: Columbia University Press, 145–231.

Hunter, L. (1991). "The Contribution of Nonhomothetic Preferences to Trade." *Journal of International Economics* 30: 345–358.

Keesing, D. B. (1965). "Labor Skills and International Trade: Evaluating Many Trade Flows with a Single Measuring Device." *Review of Economics and Statistics* 47: 287–294.

Leamer, E. E. (1980). "The Leontief Paradox, Reconsidered." *Journal of Political Economy* 88: 495–503.

———.(1984). *Sources of International Comparative Advantage: Theory and Evidence.* Cambridge: MIT Press.

———.(1988). "Measures of Openness." In R.E. Baldwin, ed., *Trade Policy Issues and Empirical Analysis.* Chicago: University of Chicago Press, 147–200.

Leontief, W. W. (1953). "Domestic Production and Foreign Trade: The American Capital Position Re-examined." *Proceedings of the American Philosophical Society* 20: 332–349.

MacDougall, G. D. A. (1951). "British and American Exports: A Study Suggested by the Theory of Comparative Costs, Part I." *Economic Journal* 61: 697–724.

Maskus, K. E. (1985). "A Test of the Heckscher-Ohlin-Vanek Theorem: The Leontief Commonplace." *Journal of International Economics* 19: 201–212.

———(1991). "Comparing International Trade Data and Product and National Characteristics Data for the Analysis of Trade Models." In P. Hooper and J. D. Richardson, eds., *International Economic Transactions: Issues in Measurement and Empirical Research.* Chicago: University of Chicago Press, 17–56.

Melvin, J. R. (1968). "Production and Trade with Two Factors and Three Goods." *American Economic Review* 63: 1249–1268.

Richardson, J. D. (1989). "Empirical Research on Trade Liberalization with Imperfect Competition: a Survey." *OECD Economic Studies* 12: 47–65.

Scherer, F. M. (1980). *Industrial Market Structure and Economic Performance.* Chicago: Rand McNally.

Stern, R. M. (1962). "British and American Productivity and Comparative Costs in International Trade." *Oxford Economic Papers* 14: 275–304.

Stern, R. M., and Maskus, K. E. (1981). "Determinants of the Structure of U.S. Foreign Trade, 1958-76." *Journal of International Economics* 11: 207–224.

Tybout, J. R. (1993). "Internal Returns to Scale as a Source of Comparative Advantage: The Evidence." *American Economic Review: Papers and Proceedings,* 83: 440–444.

Vanek, J. (1963). *The Natural Resource Content of United States Foreign Trade, 1870–1955.* Cambridge: MIT Press.

———.(1968). "The Factor Proportions Theory: The n–Factor Case." *Kyklos* 4: 749–756.

PART III

TRADE POLICY

CHAPTER 15

TARIFFS

15.1 INTRODUCTION

In Chapter 5 we contrasted a free trade equilibrium with an autarky equilibrium, in which a country does not trade at all. Both these extremes are virtually unheard of in practice. Instead, when a country does engage in trade, the government of that country will erect various barriers to restrict trade. The most common of these barriers are taxes levied on the importation of foreign goods. These taxes, commonly referred to as *tariffs,* are simply a form of commodity taxation. Tariffs are sometimes levied on exports as well as on imports. This is the case, for example, with Canadian exports of natural gas to the United States. While there are other forms of trade restriction, this chapter will concentrate on tariffs. Other barriers, such as import quotas, will be discussed in the following chapter, and we will defer until Chapter 19 a detailed discussion of why such policies are put into use by governments.

For now, we establish the two essential reasons that governments may choose to levy taxes on trade. The more important objective is to protect the operations of domestic industries that compete with imports. Taking a Heckscher-Ohlin framework, for example, we would expect import restrictions to be more severe in sectors that intensively use an economy's scarce factors. The most extreme form of protective tariff would be a tax

that eliminates imports. We refer to this as a *prohibitive tariff.* The second objective would be to raise revenues for the government. This practice is common in many developing countries, where it is easier to tax international trade at the border than to establish broad-based income taxes. Indeed, many exporters of primary products tax their foreign sales for revenue purposes. However, trade taxes are relatively unimportant as sources of revenue for developed economies.[1]

15.2 THE WELFARE LOSS FROM TARIFFS

In this section we focus on a small economy facing fixed world prices. That is, the country can trade as little or as much as it wants at a fixed world price ratio p^*. In this case, tariffs will affect the equilibrium price ratio facing domestic producers and consumers, but they will not affect p^*. Assume also that the pattern of comparative advantage is such that this country exports Y and imports X. Its government then places an *ad valorem* tax of t on each unit of X imported into the country.[2] Because p^* is fixed, the domestic price of X will rise by the full amount of the tax. Let $p = p_x/p_y$ be the domestic price ratio. Because exports are not taxed, the domestic and world prices will be related by $p_x = p_x^*(1+t)$ and $p_y = p_y^*$ or $p = p^*(1+t)$. Because of the import tariff on X, the domestic price ratio will be greater than the world price ratio ($p > p^*$).

It is tariff-distorted domestic prices, rather than world prices, that consumers pay and producers receive. Of course, trade must still be balanced at the world price ratio, because p^* remains the price ratio at which the country does business with the rest of the world. These facts give us equilibrium conditions that can be summarized as follows:

$$\text{MRS} = \text{MRT} = p = p^*(1+t) > p^* \qquad (15.1)$$

$$p_x^*(X_c - X_p) + p_y^*(Y_c - Y_p) = 0 \quad \text{or} \quad p^* = \frac{(Y_c - Y_p)}{(X_p - X_c)} \qquad (15.2)$$

where subscripts p and c again denote the amounts of a good produced and consumed, respectively. Equation (15.1) notes that domestic consumers and producers will equate the domestic MRS in consumption and MRT in production to the domestic price ratio, which is in turn greater than the world price ratio. Thus, at the post-tariff equilibrium, the slopes of the community indifference curve and the production frontier will be equal to each other but greater than the slope of the world price ratio. Equation (15.2) requires that the domestic production and consumption points be linked by the world price ratio.

These equilibrium conditions imply that the post-tariff equilibrium must be as shown in Fig. 15.1. In that diagram, A refers to the autarky equilibrium, whereas C_f and Q_f refer to the free-trade consumption and

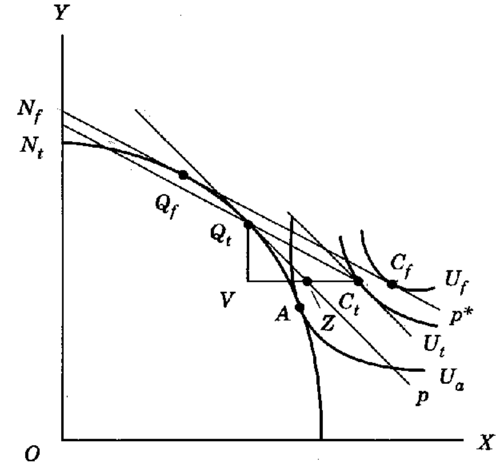


FIGURE 15.1
Effects of an import tariff.

production points, respectively. A tariff on imports of X will result in production at a point like Q_t and consumption at a point like C_t. Points Q_t and C_t are linked by the world price ratio as required by the balance-of-payments constraint (Eq. (15.2)). Points Q_t and C_t also satisfy Eq. (15.1) in that we have $\text{MRS} = \text{MRT} > p^*$.

Several characteristics of the post-tariff equilibrium are clear from Fig. 15.1. First, the post-tariff level of welfare (U_t) is lower than the free-trade level (U_f) but higher than the autarky level (U_a). Therefore, the tariff leads to a welfare loss relative to free trade but certainly not relative to autarky. Second, the tariff causes production to move from the free-trade point (Q_f) back toward the autarky point (A). Third, the reduction in imports caused by the tariff also induces a decline in the volume of exports, which must be true in the absence of any change in the world price ratio. The new trade triangle is Q_tVC_t. [redacted] Finally, because exports of VQ_t units of Y are worth VZ units of X at domestic prices but VC_t at world prices, the quantity ZC_t depicts the tariff revenue, measured in units of the import good X. We implicitly assume that the government rebates this revenue to citizens in a lump-sum fashion, allowing them to reach the consumption equilibrium at point C_t.

These effects on welfare, production, and trade reveal the essential effect of tariffs, which is to move the country back from free trade in the direction of autarky. The country specializes less in the good in which it has a comparative advantage and thus sacrifices some of the gains from trade. Indeed, real national income is reduced from ON_f to ON_t. If the tariff

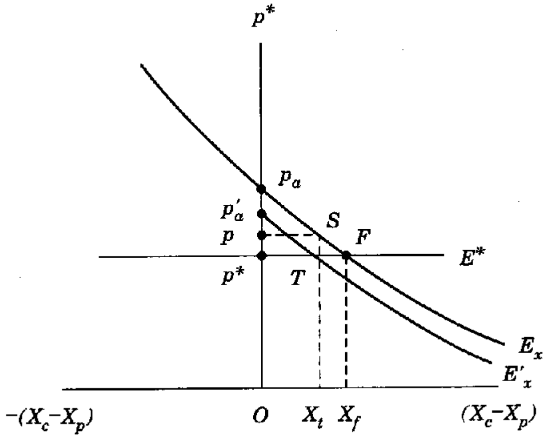


FIGURE 15.2
Effects of a tariff on excess demand curves.

were continually raised, the country would eventually find it unprofitable to import any X in Fig. 15.1 and would be driven all the way back to the autarky equilibrium at A. As we mentioned, such a tariff is called a prohibitive tariff.

Tariffs accomplish this movement toward autarky by distorting domestic prices and, because domestic producers and consumers respond to domestic prices, by distorting domestic decision making. By raising the price of X, the tariff makes it seem that X is more valuable than it actually is and thereby encourages domestic producers to produce more of it. Resources are diverted from the true pattern of comparative advantage by this misrepresentation, so gains from specialization are lost. Consumer prices are similarly distorted, so gains from exchange are also lost.

Now examine the effects of tariffs by using an excess demand curve like those developed in Chapter 4. In Fig. 15.2, the excess demand curve for the small country crosses the price axis at p_a, showing that at lower relative prices, the economy would choose to import good X. The fact that this is a small economy is depicted by the existence of a perfectly elastic foreign excess demand curve E^* at free-trade price ratio p^*. The free-trade equilibrium involves a level of imports equal to X_f at that price. An ad valorem import tariff on X imposed at rate t will shift down the import demand section of good X's excess demand curve by t percent. That is, in Figure 15.2, E'_x is defined by the relationship $p'(1+t) = p$, where p gives price along the original excess demand curve. Note that this tariff would be prohibitive if the world price p^* were between p'_a and p_a. At the world price of p^* in Fig. 15.2, the tariff reduces imports from X_f to X_t, with exports correspondingly reduced to p^*X_t units of good Y, whether the tax is imposed on imports or exports. The domestic relative price ratio in the small importer becomes $p = p^*(1+t)$, while tariff revenues are the rectangle pp^*TS, which is measured in units of Y.

An additional important point is that in addition to reducing overall income, tariffs redistribute income. In Fig. 15.1, the tariff raises the domestic price of X and shifts production from Q_f to Q_t. We know from our earlier analysis that this shift will generally change factor prices. In the Heckscher-Ohlin model, the increase in the price and production of X will increase the real income of the factor used intensively in the production of that good and decrease the real income of the other factor (the Stolper-Samuelson theorem). Thus, in this case, the losses shown in Fig. 15.1 are shared unevenly, and one factor is actually made better off. Because overall welfare in the economy falls with the tariff, it follows that the other factor suffers a welfare loss in excess of the entire welfare loss. These distributional consequences of protection help to explain why we have protectionist policies. Chapter 19 will focus on this issue.

15.3 TARIFFS, TAXES, AND DISTORTIONS[3]

As we noted earlier, tariffs are simply a special kind of tax. The purpose of this section is to expand on the notion of tariffs as taxes and to analyze the relationship of tariffs to other types of taxes. Recall from Fig. 15.1 that an import tariff on X has the effect of raising both the price charged to the consumers and the price received by the producers. This hurts the consumer of X and helps its producers. The tariff acts like a tax on consumers and a subsidy to producers. In fact, a tariff has the equivalent effect of a consumption tax combined with a production subsidy. From the information in Fig. 15.1 alone, we would find it impossible to tell whether the equilibrium at C_t was caused by an import tariff or by a combined consumption-tax/production-subsidy on X.

Import Tariffs and Export Taxes

It is somewhat more difficult to grasp the point that an import tariff on X is exactly equivalent to some export tax on Y.[4] As we indicated earlier, restricting imports is equivalent to restricting exports. Recall that an import tariff on X raises the domestic price of X above the world price ($p_x > p_x^*$) while leaving the domestic price of Y equal to the world price ($p_y = p_y^*$). The effect of the tariff on relative prices is to set $p > p^*$. For its part, an export tax establishes the following relationship between the domestic price of Y and the world price: $p_y = p_y^*(1-t)$. (Note that for this small country, the tax would reduce the domestic price received by producers

of the export good by the full amount of the tax, because exports must be sold at the fixed world price.) Thus, the tax drives a wedge between the domestic and world prices of $Y (p_y < p_y^*)$ while leaving the domestic price of X equal to the world price ($p_x = p_x^*$). The effect of the export tax is also to set $p > p^*$. [redacted]. Again looking only at the information in Fig. 15.1, we would find it impossible to tell whether the equilibrium at C_t was generated by an import tariff or by an export tax.

Export taxes and import tariffs are equivalent in that they tend to raise the relative domestic price of imports and lower the relative domestic price of exports. Both tend to shift resources out of export industries into import-competing industries. Many observers have argued that countries should restrict imports by tariffs and simultaneously encourage exports by subsidies. This contention is wrong on two accounts. First, we should not do either in a small economy, where free trade is optimal, and second, the two proposed policies have exactly opposite effects and would therefore tend to cancel each other.

Export Subsidies

It is interesting to consider the effects of subsidizing exports in a small open economy. This policy is analyzed in Fig. 15.3. Suppose that s is an ad valorem subsidy rate on exports of Y. Then $p_y = p_y^*(1 + s)$ and $p = p^*/(1 + s) < p^*$. Equation (15.1) must be replaced by

$$\text{MRS} = \text{MRT} = p = \frac{p^*}{1+s} < p^* \qquad (15.3)$$

The balance-of-payments constraint in Eq. (15.2) must, of course, continue to hold.

Figure 15.3 shows that the export subsidy causes the country to produce more Y and less X (point Q_s) than at the free trade equilibrium (point Q_f). Thus, real national income of ON_s is smaller than in free trade, as was the case with the tariff. Both policies distort resource allocation. The difference is that the subsidy generates excessive production of Y, while the tariff generates excessive production of X. Consumption occurs at C_s where the MRS in consumption equals the distorted domestic price ratio. The country trades more with the rest of the world (both exports and imports increase), but welfare is reduced from U_f to U_s. Observe the implication that artificially increasing trade by subsidizing exports will not make an economy better off in general. Indeed, export subsidies are usually more welfare-decreasing than tariffs because they require taxpayers to fund them, rather than generating tax revenues.

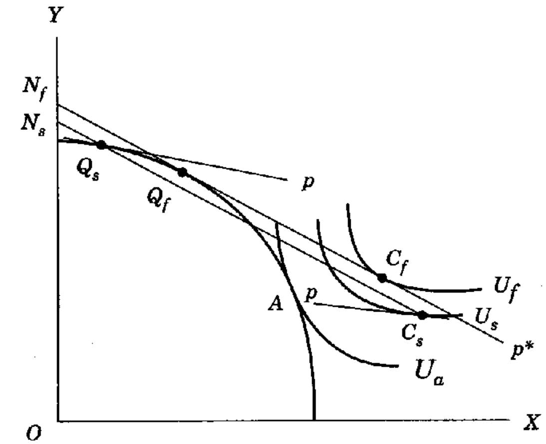


FIGURE 15.3
Export subsidies.

In fact, an export (or import) subsidy could actually make an economy worse off than in autarky, as we prove later in Section 15.7. It is possible to see this, however, by considering Fig. 15.3 again. If the production distortion induced by the subsidy is so large that the world price line emanating from point Q_s actually passes below indifference curve U_a, the country would be better off not trading than subsidizing exports.

Consumption Taxes and Production Subsidies

Now let us return to the notion that an import tariff (or export tax) is equivalent to a consumption tax and a production subsidy. Suppose that for some political reason the government is determined to increase production in the import-competing sector relative to the level it attains in free trade. One reason the government might wish to do this is that some minimum level of production in the import-competing sector is viewed as important for national security reasons, as might be the case with steel, oil, or semiconductors. Given this objective, the important economic question is, what is the least-cost method of achieving it? The problem with an import tariff is that it acts as a tax on consumption, in addition to serving as a subsidy to production. Might it not be better to use a direct output subsidy instead? The answer is definitely yes, as is shown in Fig. 15.4. If the government uses an import tariff to shift production from Q_f to Q_t, consumption will move to C_t, resulting in a welfare level of U_t.

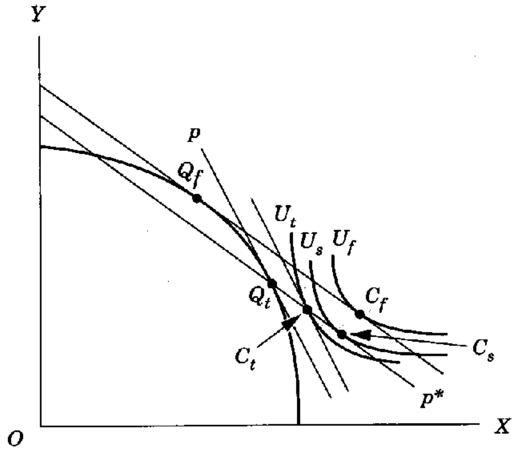


FIGURE 15.4
Consumption taxes and production subsidies.

Suppose instead that the government simply subsidizes production of X in such a way that the producer receipts per unit are the same as with the tariff. In this case output will still shift to Q_t, but consumers will not face distorted prices and will instead be allowed to trade at world prices. This will allow consumers to attain consumption bundle C_s and utility level U_s in Fig. 15.4. This result can be explained by using the terminology of gains from exchange and gains from specialization developed in Chapter 5. The tariff in Fig. 15.4 distorts both consumer and producer prices, thereby causing a loss of gains from exchange as well as gains from specialization. The subsidy distorts only producer prices and thereby causes a loss only of gains from specialization. Curiously, despite the logic in Fig. 15.4, politicians and the general public appear to find tariffs more acceptable than subsidies because tariffs are the more common method of protection. More will be said about the politics underlying such choices in Chapter 19.

Tariffs and Distortions

A final point concerning tariffs and taxes is explored in Fig. 15.5. As we noted earlier in this chapter, the result that tariffs are harmful for a small open economy relies on the assumption that there are no distortions in the economy. If there are distortions, it may be the case that tariffs could be used to offset these distortions and thereby increase welfare. This possibility is an application of what is known in economics as the *theory*

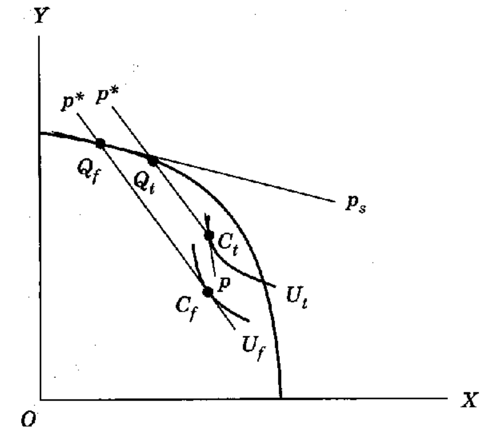


FIGURE 15.5
Tariffs and distortions.

of the second best. This theory states that in the presence of multiple distortions (such as domestic taxes or monopoly), welfare is not necessarily improved by removing a single distortion (such as an import tariff). An equivalent statement is that in the presence of distortions, adding an additional distortion may improve welfare.

An application of the latter form of the theory of the second best is shown in Fig. 15.5. Suppose that for some political reason, the producers of Y have managed to obtain a subsidy from the government and that the government is unwilling to take the political risk of removing the subsidy. Free trade production will take place at a point like Q_f in Fig. 15.5, where the domestic producer price ratio p_s (the slope of the production frontier) is flatter than the world price ratio. Consumers can trade at world prices, and so consumption is given by point C_f. Even though the government cannot remove the subsidy, it can improve welfare by introducing an additional distortion, namely, an import tariff on X. This will raise the domestic price of X and, with the rate of subsidy on Y unchanged, encourage the production of X, moving the output mix from Q_f to Q_t. The consumer prices will then be distorted by the tariff on $X (p > p^*)$, so consumption will occur at a point like C_t. Welfare is thus improved by the tariff, even though the country trades at fixed world prices. The tariff accomplishes this result by influencing production in a direction opposite to the influence of the distortionary subsidy. The effect of the tariff is to push the economy back in the direction of its efficient pattern of specialization. Of course, as our earlier analysis pointed out, imposition of the tariff could also lower welfare if the tariff-inclusive domestic price ratio does not induce much change in production but substantially worsens the consumption distortion. Such an

equilibrium would occur on an indifference curve below the one giving the original consumption choice.[5]

15.4 MONOPOLY POWER

Thus far we have assumed that the country is small and faces fixed world prices (i.e., the country is essentially a perfect competitor on world markets). Suppose now that the country is large enough that world prices will be influenced by what the country wishes to buy and sell. More specifically, the world price of our export good will fall as we export more, and the world price of our import good will rise as we import more. The more we wish to trade, the worse our terms of trade become.[6]

Figure 15.6 depicts two countries in a trading situation in which E_x^h is the home country's excess demand curve for good X and E_x^f is the foreign country's excess demand for X. An equilibrium would be established in free trade at p_f^* (see point F). Home imports are X_h^0, which coincide with foreign exports of X_f^0. Notice the effect, however, when the home country imposes a tariff (either as an import tariff on X or an export tariff on Y). The resulting downward shift in the home country's excess demand curve to $E_x^{h'}$ causes the equilibrium world price ratio to fall to p_t^* (point T) at the same time that it drives up the domestic price ratio in the home country to p (point S; recall that $p = p_t^*(1 + t)$). The resulting restriction of imports in the home country to level X_h^1 (and of foreign exports to X_f^1) is a move toward autarky. This comes about because of the higher domestic price that distorts

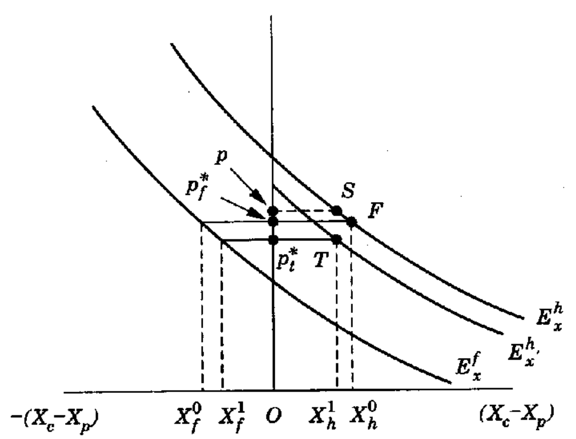


FIGURE 15.6
Terms of trade effects from tariffs.

production and consumption decisions and that would lower welfare in the home country. However, the fall in the relative equilibrium world price of the home country's import good from p_f^* to p_t^* represents a gain in the home country's terms of trade. The welfare benefits from this would, to some extent, offset the welfare losses from reduced trade. Tariff revenue is the area pp_t^*TS.

Figure 15.7 illustrates the possibility that the improvement in the terms of trade is so strong that Country H is actually made better off by the tariff. The tariff lowers the world price ratio from p_f^* to p_t^*, resulting in post-tariff production and consumption at Q_t and C_t, respectively. Home welfare increases from U_f to U_t following the imposition of the tariff.

The economic explanation of this possibility is fairly simple. While a country would like its firms to behave competitively when selling at home, it would be beneficial for the country to behave as a monopolist when selling abroad. Because we have assumed that individual firms are competitive, they cannot behave in this way. Therefore, the government can act to make the country behave as a monopolist. The tariff causes the country to restrict its "output" (exports) like a monopolist and also to restrict its "demand" (imports) like a monopsonist, thereby moving prices in the country's favor. At the same time, however, this action clearly worsens the other country's economic welfare. Country F suffers both a terms-of-trade loss and distortionary losses from the fact that the altered relative price pushes resources out of its export sector and changes consumption decisions. Accordingly, we might expect F to retaliate against

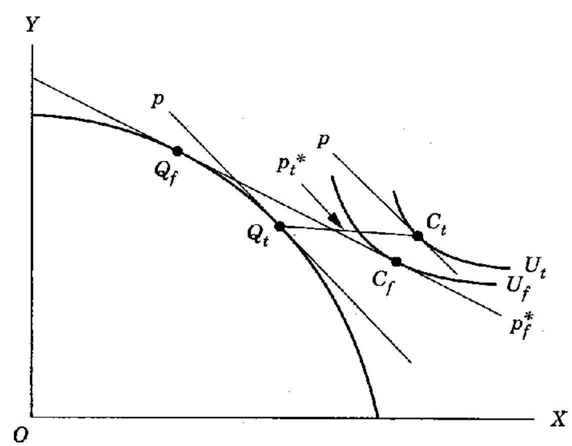


FIGURE 15.7
Welfare improvement from tariffs.

H with its own tariff on H exports. This process of tariff imposition and retaliation—a "trade war" in popular terminology—is harmful to global welfare and is likely to leave both H and F worse off than in free trade. We discuss this possibility more fully in the next section.

15.5 THE OPTIMUM TARIFF AND RETALIATION

When a country can gain by imposing a tariff, we might ask what the best possible tariff level is. This is known as the *optimal tariff* issue. The key to deriving the optimal tariff was contained in our preceding discussion, where we noted that a tariff could allow a country to exert its monopoly power in the supply of its export good or to exert its monopsony power in purchases of its import good.

Since most of our discussion has concerned tariffs on imports, let us take the latter approach and think of the country as exercising its monopsony power in the purchase of imports. Figure 15.8 depicts the situation for Country H, which is again assumed to import good X. E_x^f is the partial-equilibrium foreign supply curve (technically, the excess supply curve) and I_x^h is the home country's (excess) demand curve for X. Free trade equilibrium would be given by the intersection of the two, establishing a free trade price of p_{xf}^* and setting imports equal to I_{xf}.

But this free trade equilibrium is not a welfare optimum for Country H. An optimum requires that Country H equate the domestic price of X to

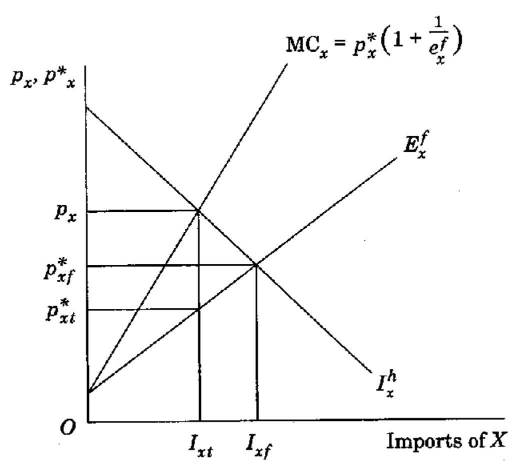


FIGURE 15.8
The optimum tariff.

the marginal cost of imports. The foreign supply curve gives the price of imports for each quantity of imports, and price is simply the average cost. Let C_x denote the total cost of imports. We have

$$C_x = p_x^* I_x; \qquad AC_x = \frac{C_x}{I_x} = p_x^*; \qquad I_x = X_c - X_p \tag{15.4}$$

where AC_x is the average cost of imports. Free trade thus equates the domestic price of X to the average cost of X ($p_x = p_x^*$).

Marginal cost is defined as the change in cost (ΔC_x) in response to a change in the quantity of imports (ΔI_x). ΔC_x can be approximated as follows.

$$\Delta C_x = p_x^* \Delta I_x + i_x \Delta p_x^* \tag{15.5}$$

Dividing Eq. (15.5) by ΔI_x, we have

$$\frac{\Delta C_x}{\Delta I_x} = p_x^* + I_x \frac{\Delta p_x^*}{\Delta I_x} = p_x^*\left[1 + \frac{I_x}{p_x^*}\frac{\Delta p_x^*}{\Delta I_x}\right] \tag{15.6}$$

Let e_x^f be defined as Country F's elasticity of supply of exports. Since $\Delta C_x / \Delta I_x = MC_x$, Eq. (15.6) can be rewritten as

$$MC_x = p_x^*\left(1 + \frac{1}{e_x^f}\right); \qquad e_x^f = \frac{p_x^*}{I_x}\frac{\Delta I_x}{\Delta p_x^*} \tag{15.7}$$

Since the supply curve slopes upward, e_x^f is positive, and marginal cost is greater than average cost ($MC_x > AC_x = p_x^*$). The marginal cost curve is shown by MC_x in Fig. 15.8. The economic reason that marginal cost of imports exceeds average cost is straightforward. Let the home country choose to import another unit of X, which would cost p_{xf}^* at the free-trade price. However, because the home country is large, its decision to import another unit would raise the world price, increasing the cost on *all* units imported. Thus, full marginal costs consist of the price of another unit of imports plus the extra inframarginal costs generated on existing imports.

Country H maximizes its welfare by equating p_x to MC_x. Thus in Fig. 15.8 the country should import quantity I_{xt}, requiring a domestic price of p_x. This restriction on imports drives down the world price of X from p_{xf}^* to p_{xt}^* and results in a welfare gain for Country H, as discussed in the previous section. We also know that an import tariff will lead to the relationship $p_x = p_x^*(1 + t)$, where t is the ad valorem tariff rate. Thus, the optimal tariff rate will be the rate that solves the equation

$$p_x = p_x^*(1 + t) = p_x^*\left(1 + \frac{1}{e_x^f}\right); \qquad t = \frac{1}{e_x^f} \tag{15.8}$$

The optimum tariff is thus equal to the inverse elasticity of foreign export supply. The more inelastic foreign excess supply is, the larger the optimum tariff will thus be (i.e., t will be large when e_x^f is small).

It is easy to see that the optimum tariff formula covers the special case of a small open economy. If a country can trade as much or as little as it wishes at fixed world prices, this means that the supply curve facing the country is horizontal, or infinitely elastic. In this situation, e_x^f is infinite and $1/e_x^f$ is zero. Thus, the optimal tariff for a small economy is zero, and free trade is indeed the optimal policy, as discussed in section 15.2.

Two qualifications of this discussion should be noted. First, the optimal tariff formula is deceptively simple in that it seems to tell us exactly what the tariff should be. But in fact, e_x^f is generally a variable that changes its value as we move along the foreign excess supply curve. The equation $t = 1/e_x^f$ is unfortunately no more than an equilibrium condition and does not by itself tell us what the numerical value of t should be. The optimal value of t must be found by first estimating E_x^f and I_x^h and then using E_x^f to construct MC_x or e_x^f. It is only after this has been done that the *formula* can be applied to find the optimal *value* of t. Note especially in this regard that t will depend on domestic factors, even though the optimal tariff formula seems to rely only on the foreign elasticity. The optimal tariff t will equal $1/e_x^f$ evaluated at the point where the domestic excess demand curve crosses MC_x. Because e_x^f generally varies along MC_x, the actual value of e_x^f and therefore t will depend on where I_x^h crosses MC_x. Thus, domestic factors do indeed help determine the value of the optimum tariff, and the simple formula does not, in fact, provide a shortcut to finding this value.

The second qualification is a much more important one. The optimal tariff discussed here is based on the assumption that the government in Country F will not retaliate when Country H institutes the tariff. However, H's optimal tariff clearly reduces F's welfare by reducing F's trade volume and worsening terms of trade (that is, it reduces the relative price of F's export good). Suppose, then, that the foreign government will exactly match any tariff that H imposes with a tariff on H's exports. From our earlier analysis we know that this tariff is equivalent to Country F imposing a tax on its exports. Thus, in Fig. 15.6, this retaliatory move will cause the export-supply portion of E_x^f (i.e., the portion to the left of the vertical axis) to shift up, which will further restrict trade and move the terms of trade back against Country H. Indeed, it is quite possible to end up with exactly the same world price ratio that prevailed with free trade. The volume of trade will be lower, and therefore, since neither country has succeeded in improving its terms of trade, both countries will be unambiguously worse off.[7]

Because of the high probability of foreign retaliation, the term "optimal tariff" is very misleading. It is optimal only under the special assumption of no retaliation. If all countries pursue this so-called "optimal" strategy simultaneously, it is likely that every country will be worse off—hardly an optimal outcome. If countries cooperate instead of myopically pursuing their self-interests, they may find that relatively free trade constitutes the optimal policy.

15.6 EFFECTIVE PROTECTION

One fundamental implication of our discussion thus far is that a tariff provides protection from imports, allowing expanded domestic production of the protected commodity. This prediction assumes that the tariff is the only tax that directly affects costs and prices of the good in question. Such an assumption is reasonable for goods that are produced solely by untraded primary inputs, such as capital and labor, and for goods that require intermediate inputs that are freely traded internationally. However, most commodities are produced with the use of intermediate goods that are themselves subject to trade taxes. Thus, a tariff on imported steel, for example, would raise costs and lower output in the automobile sector even if there were a protective tariff on cars. In general, manufacturers are better off as tariffs rise on imports that compete with their outputs and worse off as tariffs rise on their imported inputs. The term *effective protection* refers to the fact that all such tariffs need to be taken into account in computing the net protective effect of the tariff structure.

Because our goal is to isolate the costs of intermediate inputs, the notion of effective protection actually refers to the positive or negative stimulus to *value added* in production of a commodity. Value added per unit of output, v, is the difference between the price of a final good and the cost of purchasing intermediate inputs. As such, it measures the portion of the value of output that is available for payments to primary inputs. For example, if the price of an automobile is $15,000 and the cost of acquiring the steel, leather, glass, rubber, and so on needed to produce the car is $10,000, there remain $5,000 to pay for wages, the costs of capital (e.g., profits and interest), and the costs of land (e.g., rents). Value added thus captures the costs of primary inputs. In this case, value added makes up 33 percent of the gross value of the car. If the tariff structure combines to expand value added relative to free trade, it effectively raises payments to these primary factors.

In this sense, the effective rate of protection, t_e, is defined as the percentage change in a sector's value added per unit of output to a situation in which tariffs have been, v', in moving from free trade with no tariffs imposed.

$$t_e = \frac{1}{v}(v' - v) \qquad (15.9)$$

Continuing with this example, suppose for simplicity that steel is the only intermediate input in cars and that the prices given exist in free trade. We choose units so that one unit of steel is required per automobile. Now suppose that a 20-percent tariff is imposed on imported cars and that the domestic producers of cars respond to the tariff by raising their price by 20 percent, to $18,000.[8] With no tariff on steel, domestic value added thus rises to $8,000 per car, and the effective rate of protection becomes $t_e = (\$8,000 - \$5,000)/\$5,000 = 0.6$, or 60 percent. Interestingly, the 20 percent *nominal* tariff on the final good raises the *effective* tariff from zero to 60 percent. Of course, this result simply reflects the fact that value added

is a portion of gross output, so the nominal tariff has a magnified effect on value added. Thus, the share of primary inputs in cars (measured at free trade prices) rises by 60 percent, the effective protection provided them.

Suppose now that a 20-percent tariff is imposed on imported steel, raising its domestic price to $12,000, while the tariff on cars remains. The effective protection afforded to cars is now $t_e = (\$6,000 - \$5,000)/\$5,000 = 0.2$, or 20 percent. In this case, the effective rate of protection equals the nominal rate. Finally, if we raise the steel tariff to 50 percent, our computation becomes $t_e = (\$3,000 - \$5,000)/\$5,000 = -0.4$, or -40 percent. Thus, it is possible for the tariff on inputs to be high enough to reduce effective protection for the final good relative to free trade. Thus, despite the existence of the 20 percent nominal tariff on automobiles, production of cars will probably fall as the sector is forced to shed labor and other primary inputs.

From these examples we can draw certain conclusions.

1. If the tariff on the output exceeds the tariff on the input, the effective rate of protection is higher than the nominal tariff.
2. If the tariffs are equal, the effective rate of protection is equal to the nominal tariff.
3. If the output tariff is lower than the tariff on the input, the effective rate of protection is less than the nominal tariff and may even be negative.[9]

A further point to note is that tariffs are not the only determinants of effective protection. Trade taxes, domestic taxes, subsidies, quotas, and other nontariff barriers on outputs and inputs must be considered. One inference we can make is that any tariff system will be a serious disadvantage to export industries, which must sell at world prices. For an export industry, input tariffs will raise costs that may not be offset by export subsidies. Thus, we can expect that a general reduction in tariffs will be a substantial impetus to exports by improving their competitive position.

While the concept of effective protection provides important insights, it should be noted that it relies on several restrictive assumptions. Perhaps the most crucial is that all production functions exhibit constant returns to scale with fixed input coefficients. The first characteristic allows us to compute value added per unit of output without regard to the actual level of output. If constant returns to scale were not assumed, computations of value added would become more complicated. For example, suppose the automobile tariff just discussed would, by itself, double domestic output of cars. Unless steel inputs also doubled, which would not happen except under constant returns, the impact on the share of primary inputs would depend on the tariff and on returns to scale. Because we wish to isolate the changes due to tariffs, it is convenient to focus on the constant-returns case. Similarly, the assumption of fixed coefficients implies that inputs are always combined in the same proportions, regardless of factor prices or output scale. In theoretical terms, this notion means that isoquants are right-angled, or that elasticities of substitution among inputs are zero. If, however, production functions allow for substitution in production, then any change in factor prices would be expected to change the production coefficients. Like the problem with returns to scale, such changes would affect observed value-added shifts, making computations more difficult.

Removing the assumption of fixed coefficients causes two major effects. First, allowing input substitution tends to reduce calculated effective rates of protection. Second, it may alter the ranking of effective rates across sectors. However, evidence suggests that such shifts in rankings are not great even when substitution elasticities as high as 2 are assumed. Thus, it seems practical to use effective rates of protection as indicators of how the full tariff structure will influence resource allocation across industries.

A further assumption is that production and trade take place in the protected industries both before and after the tariffs are imposed. This assumption is necessary to ensure that the calculated rates actually do measure changes in value added relative to free trade levels. A final assumption is that the elasticities of foreign demand for exports, foreign supply of imports, and domestic supply of nontraded inputs are infinite. These assumptions eliminate the possibility of price changes other than those associated with the imposition of tariffs. Thus, changes in the terms of trade associated with the tariff structure are ignored. This is a major limitation on studies of effective protection because accounting for finite elasticities would change the computations significantly. Unfortunately, relatively little reliable information exists on the values of the requisite elasticities. In general it has been found that relaxing the assumption of perfectly elastic supplies of domestic, nontraded inputs tends to lower the effective rates of protection.

We should also note that allowing for more complex production processes with many required inputs complicates the process of calculating effective rates of protection, although not excessively. In particular, the importance of the various intermediate inputs must be weighted in the calculation process, and the data requirements for these calculations increase in proportion to the number of inputs.[10]

We conclude this section with two notes on the practical relevance of effective rates of protection. First, developed countries often have a structure of *escalating tariffs*, meaning that raw materials are allowed to be imported largely duty-free, while processed intermediates have higher tariffs, and finished goods have yet higher import taxes. As suggested earlier, this structure means that the effective protection provided to finished products is higher than nominal rates would suggest. In general, estimates suggest that in the industrialized countries, effective rates of protection on final goods are approximately twice the nominal tariff rates. This is especially true in certain labor-intensive products, suggesting that the tariff structure may be used to achieve substantial implicit protection. Tariff escalation continues to be a significant issue of contention between developed-country importers and developing-country exporters in multilateral trade negotiations.

TABLE 15.1
Nominal and effective rates of protection in selected industries, United States, Japan, and Republic of Korea

	United States[a]		Japan[a]		Republic of Korea[b]	
Industry	NRP (%)	ERP (%)	NRP (%)	ERP (%)	NRP (%)	ERP (%)
Agriculture	1.80	1.91	18.40	21.40	72.3	85.7
Food products	4.70	10.16	25.40	50.31	11.7	−27.6
Wearing apparel	22.70	43.30	13.80	42.20	29.0	93.8
Wood products	1.70	1.72	0.30	−30.59	8.6	6.5
Chemicals	2.40	3.66	4.80	6.39	28.5	50.9
Iron and steel	3.60	6.18	2.80	4.34	12.9	31.5
Electrical machinery	4.40	6.34	4.30	6.73	26.2	44.8
Transport equipment	2.50	1.94	1.50	0.03	31.9	12.4

Source: A. V. Deardorff and R. M. Stern (1984) and J. Yoo, (1993, 22).

[a] Nominal tariff rates and effective rates of protection to be phased in by 1986 following the Tokyo Round of tariff negotiations.

[b] Nominal tariff rates and effective rates of protection in 1982.

A second observation is that many developing countries have arranged their protective structures so that effective tariffs are far higher than published tariffs. In part, this is an attempt to foster growth in domestic manufacturing through a regime of *import-substituting industrialization.* Again, the goal is to promote domestic output of final goods by escalating tariffs on the inputs. This policy is often accompanied by a deliberate overvaluation of the domestic currency, done in part to discourage exports of primary products in favor of keeping primary goods at home for use in manufacturing import-competing goods. On occasion, these protection levels can be extraordinary. For example, it was estimated that in 1969, Argentina had nominal tariff rates of 63 percent on finished textiles and 76 percent on woodworking industries. However, the associated effective rates of protection, accounting for trade barriers, taxes, and the exchange-rate regime, amounted to 832 percent and 1,308 percent, respectively.[11]

To provide further perspective, Table 15.1 presents recent nominal tariff rates and estimates of effective protection rates in the United States, Japan, and the Republic of Korea. It can be seen that all countries heavily protect wearing apparel, while Japan and Korea strongly protect their agricultural sectors. Indeed, the costs of agricultural protection in Korea are so high that the food products industry is effectively taxed, despite an 11.7 percent nominal tariff.

15.7 GAINS FROM TRADE WITH MANY GOODS, TRADE TAXES, AND SUBSIDIES

Up to this point, we have focused mainly on a simple model in which there are only two goods. Results of the analysis suggest that an import tariff reduces welfare relative to free trade but still leaves welfare greater than

or equal to the autarky level. An import or export subsidy, however, has the potential of making the country worse off relative to autarky.

Can we say anything about a country that trades many goods, some of which are taxed and some of which are subsidized? It turns out that there is in fact a very simple condition for gains from trade: if net trade tax revenue (the sum of all import and export tax revenues minus trade subsidy payments) is positive, then the country is better off than in autarky. Another way of saying this is that if trade is, on average, taxed more than it is subsidized, then there are still gains from trade.

Suppose that there are n goods, $(X_1 \ldots X_n)$, with fixed world prices $(p_1^* \ldots p_n^*)$ (although the argument easily generalizes to a large economy) and corresponding domestic prices $(p_1 \ldots p_n)$. Domestic and world prices are related by $p_i = p_i^*(1+t_i)$. If good i is imported, then a positive t_i is an import tariff and a negative t_i is an import subsidy. If good i is exported, then a positive t_i is an export subsidy and a negative t_i is an export tax (e.g., if the world price is higher than the domestic price ($t_i < 0$), then t_i is an export tax).

Domestic producers optimize with respect to domestic prices. In a competitive economy, this means that the value of production at *domestic* prices in (distorted) trade is greater than or equal to the value of autarky production at these same prices. Let superscript d denote quantities in tax/subsidy distorted trade and superscript a denote autarky quantities. Subscripts p and c denote production and consumption quantities as before. Competitive equilibrium is characterized by

$$\sum_i^n p_i^*(1+t_i)X_{ip}^d \geq \sum_i^n p_i^*(1+t_i)X_{ip}^a \tag{15.10}$$

Equation (15.10) can be rearranged as

$$\sum_i^N p_i^* X_{ip}^d \geq \sum_i^n p_i^* X_{ip}^a + \sum_i^n p_i^* t_i (X_{ip}^a - X_{ip}^d) \tag{15.11}$$

The balance-of-trade condition and the autarky market-clearing condition are given by

$$\sum_i^n p_i^* X_{ip}^d = \sum_i^n p_i^* X_{ic}^d \qquad X_{ip}^a = X_{ic}^a \tag{15.12}$$

Substituting Eq. (15.12) into Eq. (15.11), the latter becomes

$$\sum_i^n p_i^* X_{ic}^d \geq \sum_i^n p_i^* X_{ic}^a + \sum_i^n p_i^* t_i (X_{ip}^a - X_{ip}^d) \tag{15.13}$$

A welfare comparison of distorted-trade versus autarky consumption must, however, be done at domestic prices. By adding and subtracting several terms to both sides of the equation, Eq. (15.13) becomes

$$\sum_I^n p_i^*(1+t_i)X_{ic}^d \geq \sum_i^n p_i^*(1+t_i)X_{ic}^a + \sum_i^n p_i^* t_i \left[(X_{ic}^d - X_{ip}^d) - (X_{ic}^a - X_{ip}^a)\right] \tag{15.14}$$

Consider the term in square brackets on the right-hand side of Eq. (15.14). Let $M_i = X_{ic} - X_{ip}$ be the imports of good i. In autarky, these are zero. Thus, the term in square brackets on the right-hand side of Eq. (15.14) reduces to M_i^d. Using the relationship between world and domestic prices, Eq. (15.14) then reduces to a relatively simple expression.

$$\sum_i^n p_i X_{ic}^d \geq \sum_i^n p_i X_{ic}^a + \sum_i^n p_i^* t_i M_i^d \qquad (15.15)$$

The second summation on the right-hand side of Eq. (15.15) gives the total net value of trade tax revenue. Equation (15.15) gives us the welfare comparison we are seeking. The value of domestic consumption in the distorted trade equilibrium is greater than or equal to the value of autarky consumption (evaluated at distorted-trade, domestic prices) if net trade tax revenue is positive. Again, recall that if X_i is imported ($M_i > 0$), $t_i > 0$ is an import tariff, and if X_i is exported ($M_i < 0$), $t_i < 0$ is an export tax. A country whose taxes on trade exceed its subsidies cannot be worse off than in autarky, but a country whose subsidies exceed its taxes on trade can be worse off.

15.8 CONCLUDING REMARKS

Tariffs, which are taxes on imports, are the most common form of government interference with international trade. They exist largely to protect domestic firms that compete with imports, though tariffs are sometimes levied for purposes of raising government revenues. We have analyzed tariffs and other trade taxes and subsidies in our general equilibrium framework, making the following major points.

1. Tariffs, like other forms of commodity taxes, alter relative prices and change quantities transacted. Import tariffs raise the domestic price of imports and so reduce their quantity. If a country is small and thus faces fixed world prices, this reduction causes a reduction in the gains from trade, thereby reducing national welfare. Tariffs simply move the country back in the direction of autarky.
2. The effects of tariffs, as policy distortions, on consumers and producers can be examined in terms of both the import-competing and export sectors. Specifically, an import tariff is equivalent to a combination of a tax on consumption of the import good combined with a subsidy on production of the import-competing good. Furthermore, an import tax is equivalent to an export tax (in a two-sector model), and an import tariff can be offset by an export subsidy.
3. While tariffs are almost surely welfare-reducing for a small country,[12] a large country may be able to exploit its monopoly power on world markets by using a tariff to gain a favorable terms-of-trade effect for itself. However, this power is greatly undermined by the likelihood that other countries will retaliate against the country imposing the tariff. The noncooperative scenario of tariff imposition and retaliation is unlikely to benefit any country. The probable outcome is that the volume of world trade will be reduced without a significant gain for any country in terms-of-trade advantage. Gains from trade are lost with the reduction in world trade. Stated in other terms, all countries suffer efficiency losses in terms of distorted consumption and production decisions with no offsetting advantages in international relative prices. However, in a cooperative scenario, countries are likely to find that the so-called "optimal" tariffs are not, in fact, optimal at all. By cooperatively reducing their tariffs, countries can expand trade without suffering adverse terms-of-trade effects and can thereby increase their welfare.
4. Tariffs frequently fall upon imported intermediate goods in the production process. The effect of these tariffs is to generate a pattern of distortions in the use of intermediate goods in the production process. The real extent of tariff protection in such cases is hidden by the layering of tariffs, and effective protection on final goods may be much greater or smaller than nominal tariff rates would indicate. It is important to account for these intersectoral effects of the tax structure in order to understand the nature of resource allocation pressures.

PROBLEMS

1. Is it possible for a tariff to make a country worse off than in free trade?
2. Redraw Fig. 15.1 for the case in which the country exports X and imports Y.
3. Suppose a country produces two goods, X and Y, along a linear production transformation curve where X is the imported good and production is specialized in Y. Show the welfare losses from the imposition of a tariff on X and decompose these losses into production losses and consumption losses.
4. The government of the Republic of Korea at one time followed an aggressive policy of import protection with tariffs and export promotion with subsidies. Contrast this policy with one of free trade.
5. For the past several years, the United States has been subsidizing exports of grain. Analyze the welfare effects of this policy.
6. Suppose that instead of instituting an optimal import tariff, the government decides to use an optimal export tax. Using a methodology similar to that of Fig. 15.8, can you derive the optimal export tax formula? (Hint: in this case the home country would act as a monopolist in its export good).
7. Suppose a commodity is produced only after the introduction of a tariff on imports with which it competes. Will the usual effective-protection calculation be correct in this case?
8. Assume that yarn is the only input into cloth and that the proportion of yarn in the unit cost of cloth is 0.4 (that is, $a_{yc} = 0.4$). Let the nominal tariff rate on cloth be 25%. Compute the effective rates of protection to cloth when the yarn tariff is: (a) 0%, (b) 10%, (c) 25%, (d) 50%. Solve for the tariff rate on yarn that would make the effective rate on cloth zero.

9. Consider an industry X_1 with two intermediate inputs, X_{21} and X_{31}, and suppose that $a_{21} = 0.2$ and $a_{31} = 0.5$. Tariff rates are $t_1 = 30\%$, $t_2 = 20\%$, and $t_3 = 10\%$. Calculate the effective rate of protection on X_1.

10. Suppose that a small country exports both goods X_1 and X_{21}, where X_{21} is an input into X_1. Let $a_{21} = 0.3$ and suppose that the export tax rates are $t_1 = 25\%$ and $t_2 = 10\%$. Can you compute the effective rate of tax on X_1?

NOTES

1. For example, tariff revenue amounts to only 0.01 percent of total government revenue in the United Kingdom, 0.02 percent in Germany, and 1.56 percent in the United States. On the other hand, tariffs provide the government of Argentina with 13.31 percent of total revenue and the government of Ghana with 40.90 percent of all revenues. The source for these figures is International Monetary Fund (1986).
2. An *ad valorem* tax is one that is levied as a percentage of the value of the taxed goods, like a sales or excise tax. Not all tariffs are of this form. Some tariffs are specific tariffs, levied per unit of quantity, as in so many dollars per ton of fish. Frequently, specific and ad valorem tariffs are combined into compound tariffs.
3. A comprehensive treatment of the effects of various taxes on trade and welfare is given by Melvin (1975).
4. This point is credited to Lerner (1936), who assumed competitive markets in its proof. See also Kaempfer and Tower (1982) for extensions of numerous complications to the simple model.
5. Certain aspects of trade policy in the presence of domestic distortions are discussed in Bhagwati (1967, 1971), Johnson (1965), and Melvin (1975).
6. The original argument was made by Bickerdike (1906) and was formalized and generalized by a number of authors, including Graaff (1949).
7. See Johnson (1954) for a full discussion of the tariff war outcome and Markusen (1981) for an analysis of the opposite problem of cooperative tariff reduction. Note that the *threat* of retaliation by F might be sufficient to deter the initial decision by H to impose its optimal tariff, so there is a strategic element to this analysis as well.
8. Thus, we assume that the price of cars goes up by 20 percent, whether they are produced domestically or imported. Implicitly, this means that the importing country is small and that imported and domestic cars are perfect substitutes. We retain these assumptions with respect to intermediate inputs as well.
9. A further possibility is the existence of *negative value added,* which refers to the unusual situation in which an inefficient domestic activity is so heavily protected that it is induced to produce a certain output, despite the fact that at world prices, the value of that output is less than the value of the intermediate inputs. For example, heavily distorted economies produce some products, such as cars, that cannot be sold on world markets for a price that would cover input costs.
10. In the presence of multiple inputs into the production of some good j, the value added of j with protection on inputs will be

$$v_j' = (1 + t_j) - \sum_{i=1}^{n} a_{ij}(1 + t_i)$$

Here, a_{ij} denotes input-output coefficients, and t refers to the tariff rate on the output and on the inputs. To derive value added in free trade, this equation can be recalculated with all tariff rates set to zero. The resulting v_j' and v_j can then be substituted into Eq. (15.9) in order to calculate the effective rate.
11. See the paper by J. Berlinski and D. M. Schydlowsky (1982).
12. Recall the possibility that a tariff could raise welfare in the face of some other domestic distortion.

REFERENCES

Berlinski, J., and Schydlowsky, D. M. (1982). "Argentina." In B. Balassa, ed., *Development Strategies in Semi-industrial Economies,* Washington, DC: The World Bank.

Bhagwati, J. N. (1967). "Non-Economic Objectives and the Efficiency Properties of Trade." *Journal of Political Economy* 75: 738–742.

———. (1971). "The Generalized Theory of Distortions and Welfare." In J. Bhagwati et al., eds., *Trade, Balance of Payments and Growth: Essays in Honor of Charles P. Kindleberger,* Amsterdam: North-Holland.

Bickerdike, D. F. (1906). "The Theory of Incipient Taxes." *Economic Journal* 16: 529–535.

Corden, W. M. (1974). *Trade Policy and Economic Welfare.* Oxford: Clarendon Press.

Deardorff, A. V., and Stern, R. M. (1984). "The Effects of the Tokyo Round on the Structure of Protection." In R. E. Baldwin and A. O. Krueger, eds., *The Structure and Evolution of Recent U.S. Trade Policy.* Chicago: University of Chicago Press.

Graaff, J. de V. (1949). "On Optimal Tariff Structures." *Review of Economic Studies* 16: 47–59.

Johnson, H. G. (1954). "Optimal Tariffs and Retaliation." *Review of Economic Studies* 21: 142–153.

———. (1965). "Optimal Trade Intervention in the Presence of Domestic Distortions." In R. Caves et al., eds., *Trade, Growth and the Balance of Payments. Essays in Honor of G. Haberler.* New York: Rand McNally.

Kaempfer, W. H., and Tower, E. (1982). "The Balance of Payments Approach to Trade Tax Symmetry Theorems." *Weltwirtschafliches Archiv.* 118:148–165.

Lerner, A. (1936). "The Symmetry between Import and Export Taxes." *Economica* 11: 306–313.

Markusen, J. R. (1981). "The Distribution of Gains from Bilateral Tariff Reduction." *Journal of International Economics* 11: 553–572.

Melvin, J. R. (1975). *The Tax Structure and Canadian Trade.* Ottawa: Economic Council of Canada.

Yoo, Jung-ho. (1993). "The Political Economy of Protection Structure in Korea." In A. O. Krueger and T. Ito, eds., *Trade and Protectionism.* Chicago: University of Chicago Press.